Non-parametric calibration

Osborne, Christine

Award date:
1990

Awarding institution:
University of Bath

NON-PARAMETRIC CALIBRATION

submitted by

CHRISTINE OSBORNE

for the degree of Ph.D. of the
University of Bath
1990

C. Osborne

C. Osborne

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

This thesis is devoted to the development of several non-parametric Bayesian approaches to univariate calibration, whilst reviewing the wide variety of approaches to both univariate and multivariate calibration.

A typical calibration procedure consists of two stages (i) a calibration experiment consisting of n observations, (x_i, Y_i), which are used to estimate the true calibration curve, f, and (ii) a prediction stage where one or more observations, Y_j', are taken which correspond to an *unknown* value of X. Suppose the unknown value is denoted by ξ, then the aim of the calibration procedure is to draw inferences concerning ξ.

All the approaches developed in the thesis involve deriving the posterior distribution of ξ, given both the calibration data $\{(x_i, Y_i)\ i=1,2,3,....n\}$ and the observations Y_j'. Both point and interval estimates for ξ are obtained from this posterior distribution. A key assumption for most of the thesis is that f is a cubic spline with knots $\{x_i\}\ i=1,2,....n$.

The first two approaches to be developed view ξ as a non-linear functional of f and simulation from the posterior distribution of f is used to obtain the posterior distribution of ξ. The third approach uses predictive densities and involves an approximation to the distribution of the residual sum of squares of the calibration experiment. The fourth and fifth approaches are developed especially for data sets where the observations (Y_i, Y_j') are proportions. The fourth approach results in a non-parametric version of logistic regression and the fifth approach involves a logistic transformation of the observations (Y_i, Y_j').

All the non-parametric methods are assessed by testing them out on a wide variety of both simulated and real data sets. Finally, the first three approaches developed in the thesis are compared with each other and also with a selected number of parametric approaches. Conclusions are drawn concerning the relative merits of the non-parametric approaches.

Acknowledgements

I would like to thank the following people for their help in the production of this thesis:

Professor Bernard Silverman, my supervisor, who has guided my research and provided many helpful comments and suggestions.

My husband Norman who, apart from his proof reading expertise, has given tremendous moral support.

Mike, Paul, Shirley, and Martyn whose daily companionship has been much appreciated and Glenn for his helpful computing advice.

The Science and Engineering Research Council who funded my postgraduate studentship.

Table of Contents

Chapter Seven. COMPARISONS AND CONCLUSIONS.

1. INTRODUCTION

Statistical calibration has some similarities with scientific calibration, which is the process whereby the scale of a measuring instrument is determined or adjusted on the basis of an informative or calibration experiment, but it has a more complicated form. A statistical calibration problem is a kind of inverse prediction, a problem of retrospection and some authors call it inverse regression rather than calibration. It is probably best explained by considering a typical univariate calibration problem.

Consider the problem of a chemist wishing to establish a calibration curve to use in measuring the amount of a certain chemical A in samples sent to an analytical laboratory. There are two stages in the calibration process. In the first stage, for each of n samples with *known* amounts of chemical A, one or more measurements are made with the relatively quick, inexpensive *test method* being calibrated (Y). The *known* amounts of chemical A have been determined by an extremely accurate *standard method* that is slow and expensive (X). The resulting data constitutes the calibration experiment and is used to estimate the calibration curve f. This calibration curve is now ready for use in the second stage of the calibration process which involves prediction. In the second stage, samples with *unknown* amounts of chemical A are analysed with the test method and the amount of chemical A predicted for each new sample. For a given sample, one or more measurements using the test method may be made.

The chemist's problem outlined above illustrates a common calibration procedure. Let us consider the mathematical formulation of this calibration procedure. Let the true values associated with the standard and test methods be designated by ξ and η. As this thesis is concerned with non-parametric calibration, it will be assumed that $\eta = f(\xi)$ where f is a smooth function, in most of the chapters f being a natural cubic spline.

In the first stage of the calibration process, the *calibration experiment, n* pairs of observations (X_i, Y_i) are obtained, where X_i and Y_i are the observed values of ξ_i and η_i respectively. Now

$$\begin{aligned} X_i &= \xi_i + \delta_i \qquad i=1,2,....n \\ Y_i &= \eta_i + \varepsilon_i \qquad i=1,2,....n \end{aligned} \tag{1.1}$$

where ε_i and δ_i are experimental errors. In all *absolute* calibration problems $\delta_i = 0$. The chemist's problem outlined above is an example of absolute calibration and this thesis will be concerned with absolute calibration problems. It is usual in calibration to assume that the calibration curve is monotonic. The

model for the calibration experiment is therefore given by

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,3.....n \tag{1.2}$$

where it is assumed that f is a monotonic natural cubic spline with knots at the data points $\{ x_i \}$ $i=1,2,....n$. It is usual to assume that $\varepsilon_1 , \varepsilon_2 , \varepsilon_3 ..... \varepsilon_n$ are independent normal random variables with mean zero and variance σ_1^2. We will assume that $\sigma_1^2 = \sigma^2/w_i$ where the weights w_i are known.

Having estimated the calibration curve, we proceed to the second stage of the calibration process, the *prediction stage*. A sample is presented with a specific unknown value of η and one or more measurements (Y') are made using the test method. The model for the prediction stage is given by

$$Y_j' = \eta + \varepsilon_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3.....m \tag{1.3}$$

where the ε_j' are assumed to be independent normal random variables with mean zero and variance σ_2^2. We will assume that $\sigma_2^2 = \sigma^2$.

Given the data from the first and second stages, inferences are now made about the *unknown* ξ, which corresponds to η for the sample being measured. The value of ξ is given by

$$\xi = f^{-1}(\eta) \tag{1.4}$$

Chapters 3, 4, 5 and 6 of this thesis are concerned with different Bayesian approaches to this inference problem whereas Chapter 2 reviews the many varied approaches (Bayesian and non-Bayesian, parametric and non-parametric) to univariate and multivariate calibration problems.

In Chapters 3 and 4, ξ is viewed as a non-linear functional of the true calibration curve f and simulation is used to generate the posterior distribution of ξ. In these chapters, we simulate from the posterior distibution of f, which is a truncated multivariate normal distribution, truncated in such a way as to remove non-monotonic splines. Suppose such a posterior realisation of f is called $\tilde{f}_s$. In Chapter 3, a non-informative prior for η is assumed and posterior realisations (η_v) of η are generated. Then $\tilde{f}_s^{-1}(\eta_v)$ is evaluated a large number of times to generate the posterior distribution of ξ. However in Chapter 4, having simulated from the posterior distribution of f, a switching algorithm is used to generate the posterior distribution of ξ, given a particular prior for ξ. The method is tested out on the same data sets as those used in Chapter 3, using a variety of priors for ξ. The main difference between the two approaches in Chapters 3 and 4 is that in Chapter 4 an explicit prior for ξ is used, facilitating the use of prior information on ξ, whereas in Chapter 3 an implicit prior for ξ is used since from equation (1.4) the assumed prior for η implies a particular prior for ξ.

The approach of Chapter 5 is completely different from the approaches of Chapters 3 and 4 because it is based on predictive densities and does not use simulation. The approach requires, for all but one data set considered, an approximation to the distribution of the residual sum of squares for the calibration experiment. The methods developed in this chapter are based on the papers of Imhof (1961), Solomon and Stephens (1977) and Eagleson and Buckley (1988) and they are assessed by testing them out on a variety of data sets.

Chapter 6 considers two non-parametric approaches to data sets where the observations (Y_i, Y_j') are proportions. The first approach considers the analysis of proportion data arising from a Binomial model and results in a non-parametric version of logistic regression based on cubic smoothing splines. This is tested out on two simulated data sets. The second approach considers non-Binomial proportions and involves transforming the data. Once transformed, it is possible to apply the methods of Chapters 3, 4 and 5 to produce the posterior distribution of ξ. By way of an illustration, the method developed in Chapter 3 is applied to a teeth data set.

The first part of Chapter 7 is devoted to the comparison of the non-parametric methods described in Chapters 3, 4 and 5. The chapter then goes on to compare, using data arising from linear models, the parametric methods of Hoadley (1970), Aitchison and Dunsmore (1975) and Hunter and Lamboy (1981a) with the non-parametric methods developed in Chapters 3, 4 and 5. The final part of the chapter considers the conclusions to be drawn as a result of these comparisons and makes brief suggestions for possible future work.

2. A REVIEW OF CALIBRATION

2.1 THE NATURE OF CALIBRATION

2.1.1 Absolute and comparative calibration

It is important in any discussion of calibration to distinguish between absolute and comparative calibration. Williams (1969a) stressed that although these two activities are both called calibration, they are conceptually different and lead to different issues in statistical modelling.

In absolute calibration a quick or non-standard measurement technique is calibrated against a standard or defined measurement. The standard measurement is either known or made with negligible error.

With comparative calibration one instrument or measurement technique is calibrated against another with neither one being inherently a standard so that there is no standard measurement X. Williams (1969a), Barnett (1966,1969) and Theobald and Mallinson (1978) discuss comparative calibration and an extensive practical example of it is given by Sayers et al. (1986) who analysed the data from the International Road Roughness Experiment (IRRE) conducted in Brazil. The IRRE was proposed in order to find a standard road roughness index appropriate for the many types of roughness measuring equipment now in international use, and to provide a basis for comparing roughness measures obtained by different procedures and instruments. Cochran (1943) proposed methods of multivariate analysis for application in comparative calibration experiments where the values obtained on two different scales are correlated. Rosenblatt and Spiegelman (1981) give several references in the field of comparative calibration.

2.1.2 Mathematical formulation of the univariate calibration problem

As stated in Chapter 1, a common calibration procedure consists of two stages (i) the calibration experiment and (ii) the prediction stage.

The calibration experiment consists of n pairs of observations (X_i, Y_i). From equations (1.1)

$$X_i = \xi_i + \delta_i \qquad i=1,2,....n$$

$$Y_i = \eta_i + \varepsilon_i \qquad i=1,2,....n$$

where ε_i and δ_i are experimental errors. In absolute calibration $\delta_i = 0$ for all i.

Lwin and Spiegelman (1986) consider a case close to absolute calibration where none of the δ_i are identically zero but the δ_i have a known finite bound. Mandel (1984) considers calibration when both X and Y have error. Making the assumption that $\delta_i = 0$ for all i, produces the following model for the calibration experiment

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots n$$

where it is assumed that $\eta = f(\xi)$. There have been many papers devoted to the linear calibration problem i.e. it is assumed that $\eta = f(\xi) = \beta_0 + \beta_1\xi$ where β_0 and β_1 are the intercept and slope parameters respectively. The model for the calibration experiment in this case becomes

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i=1,2,\ldots n \tag{2.1.1}$$

Tallis (1969), Scheffé (1973), Lwin and Maritz (1980), Clark (1980), Lundberg and De Maré (1980) and Lwin and Spiegelman (1986) have assumed a general form for $f(\xi)$ to derive their theoretical results (see Sections 2.2, 2.3, 2.4 and 2.5). The majority of them consider the linear model as a special case of interest. Knafl et al. (1984) assumes a non-linear form for $f(\xi)$ in response to the particular demands of their pressure-volume calibration problem (see Section 2.5). The next assumption which is usually made is that the ε_i's are i.i.d normal random variables with mean zero and variance σ_1^2.

At the prediction stage of the calibration process the model is given by

$$Y_j' = \eta + \varepsilon_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,\ldots.m$$

With the assumption of a linear model this becomes

$$Y_j' = \beta_0 + \beta_1\xi + \varepsilon_j' \qquad j=1,2,\ldots m \tag{2.1.2}$$

where the ε_j' are assumed to be i.i.d. normal random variables with mean zero and variance σ_2^2. Berkson (1969) argued that σ_1^2 will ordinarily be substantially smaller than σ_2^2, since the calibration experiment is usually performed under relatively highly controlled conditions. Perng and Tong (1974) in their sequential approach to calibration assume that $\sigma_1^2 \neq \sigma_2^2$ but the majority of authors assume that these variances are equal, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

The objective of the absolute calibration procedure is to make inferences about the *unknown* ξ which corresponds to η for the new sample being measured at the prediction stage. The value of ξ is given by equation (1.4)

$$\xi = f^{-1}(\eta)$$

In the case of the linear model, this equation becomes

$$\xi = \frac{(\eta - \beta_0)}{\beta_1} \tag{2.1.3}$$

There are many approaches to this inference problem and these will be outlined in the sections which follow.

In comparative calibration, one is typically concerned with the calibration of p measuring instruments when each is used to make measurements of the same property on every member of a group of specimens. There is usually no standard measurement to which they may be referred so there is no question of estimating the measurement expected on one instrument corresponding to that observed on another as in absolute calibration for example. Instead one is concerned in estimating a set of symmetrical calibration equations relating the expected measurements of pairs of instruments. Theobald and Mallinson (1978) put forward the following mathematical model.

Suppose ξ_i represents the unobserved 'true' reading of the ith instrument (i=1,2,...p) and Y_i is the observed measurement of the ith instrument so that

$$Y_i = \xi_i + \varepsilon_i \qquad i=1,2,...p. \tag{2.1.4}$$

where the ε's represent the errors of measurement. Assume that ε_i is independent of ξ_i and the ε_i are independent random variables with mean zero and variance σ_i^2 (i=1,2,...p) which are constant from specimen to specimen. Suppose further that

$$\xi_i = \mu_i + \lambda_i F \qquad i=1,2,...p \tag{2.1.5}$$

where F represents a hypothetical standard measurement having zero mean and unit variance over the population of possible specimens. The λ_i are called *calibration factors*.

Substituting (2.1.5) into (2.1.4) gives the model

$$Y_i = \mu_i + \lambda_i F + \varepsilon_i \qquad i=1,2,...p \tag{2.1.6}$$

If F and ε have normal distributions, model (2.1.6) is the standard model for factor analysis restricted to a single common factor so the parameters λ_i and μ_i (i=1,2,...p) can be estimated using a program for carrying out factor analysis by maximum likelihood (ML). The calibration equation for any two instruments h and l is

$$\xi_h = \mu_h + (\xi_l - \mu_l)\frac{\lambda_h}{\lambda_l} \qquad h,l=1,2,...p.$$

When $\lambda_h = \lambda_l$, the true measurements differ by an additive bias so several authors (Grubbs, 1948,1973; and Theobald and Mallinson, 1978) have considered the

estimation of the parameters of the model (2.1.6) subject to the λ_i's all being the same. It is also possible to estimate the precision of an instrument. Let π_i represent the precision of the ith instrument where

$$\pi_i = \frac{\text{St.dev}\,(\xi_i)}{\text{St.dev}\,(\varepsilon_i)} = \frac{\lambda_i}{\sigma_i} \qquad i=1,2,\ldots p.$$

The parameters π_i can be estimated using ML estimation (Theobald and Mallinson, 1978).

Williams (1969a) uses a different model from (2.1.6). Suppose that there are n specimens to be measured by the p instruments. Let ξ_{ij} represent the true measurement of the jth specimen tested on the ith instrument. Then Williams assumes that

$$\xi_{ij} = \mu_i + \lambda_i \varphi_j \qquad i=1,2,\ldots p \quad j=1,2,\ldots n$$

where φ_j are scalar parameters rather than values taken by the random variable F as in model (2.1.6). This leads to the model

$$Y_{ij} = \mu_i + \lambda_i \varphi_j + \varepsilon_{ij} \tag{2.1.7}$$

where it is assumed that the ε_{ij}'s are independent random variables with mean zero and variance σ_i^2 $(i=1,2,\ldots p)$.

Theobald and Mallinson (1978) strongly criticise William's use of the *ML* method to estimate the parameters of model (2.1.7). They state that the *ML* method cannot be applied satisfactorily when there is no prior information on the error variances and replicate observations are not available. They suggest a consistent method of estimation for model (2.1.7).

2.1.3 Controlled and random/natural calibration

In controlled calibration the values of X are fixed or pre-chosen. The calibration experiment is often designed so that chosen values of X span the range of possible values of X. The chemist's problem described in Chapter 1 is an example of a controlled calibration problem. In random or natural calibration the X-values are not chosen by the experimenter but X is a random variable as well as Y. Often X will correspond to random true values accurately determined. Brown (1982) emphasises the importance of distinguishing between controlled and random calibration. Aitchison and Dunsmore (1975) give examples of controlled and random calibration as does Brown (1982). Naes (1985a) comments that when using Beer's model in chemical spectroscopy, the basic distinction between natural and controlled calibration is unimportant if the residuals, after fitting the calibration model to the calibration data, are small, but if appreciable errors exist,

the distinction is important (see Section 2.6).

2.1.4 Milestones.

Rosenblatt and Spiegelman (1981) offer a detailed general discussion of calibration. The papers which have acted as milestones along the development road of statistical calibration theory are Eisenhart (1939), Krutchkoff (1967), Hoadley (1970), Scheffé (1973) and Brown (1982). These will be considered in detail below.

2.2 THE CLASSICAL AND INVERSE APPROACHES TO CALIBRATION

2.2.1 The classical estimator

Eisenhart (1939) set the stage for statistical investigations of absolute calibration problems. His analysis and solution of the inverse estimation problem has come to be called *classical*. Eisenhart obtained his estimate of ξ by considering the regression of Y on X,

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

The estimated regression line of Y on X is given by

$$\begin{aligned} \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X \\ &= \bar{Y} + \frac{S_{xy}}{S_{xx}}(X-\bar{x}) \end{aligned} \qquad (2.2.1)$$

where $S_{xy} = \sum_i (x_i-\bar{x})(Y_i-\bar{Y})$ and $S_{xx} = \sum_i (x_i-\bar{x})^2$. Eisenhart then inverted equation (2.2.1) to give an estimator of ξ, the unknown X, which has since become known as the *classical estimator*. Let it be denoted by $\hat{\xi}_c$. Then

$$\hat{\xi}_c = \bar{x} + \frac{S_{xx}}{S_{xy}}(\bar{Y}'-\bar{Y})$$

where $\bar{Y}'$ is the mean of the m observations at the prediction stage. If one makes the assumption of normal errors in the models (2.1.1) and (2.1.2), then $\hat{\xi}_c$ is the maximum likelihood estimator of ξ. Eisenhart also produced an interval estimate for ξ based on the t-distribution with $(n-2)$ degrees of freedom.

Subsequent texts and journal articles used the same *classical* approach as favoured by Eisenhart, i.e. regressing Y on X. Examples are Mandel and Linnig (1957), Mandel (1958), Williams (1959) and Linnig and Mandel (1964). Mandel and Linnig (1957) adopted a joint inference approach to the problem of interval estimation of ξ and produced conservative intervals when compared with

Eisenhart's.

Fieller (1954) produced interval estimates for ξ identical to those of Eisenhart using a fiducial argument. Fieller showed that the calibration problem could be reduced to considering the ratio of the means of two normally distributed random variables (see Hoadley, 1970). Creasy (1954) also used a fiducial argument to obtain interval estimates.

The *classical* approach to interval estimation has caused considerable consternation over the years because if the slope parameter β_1 is not significantly different from zero the interval is either the whole real line or even two disjoint semi-infinite lines. As a result of this problem, Berkson (1969) and Shukla (1972) obtained asymptotic expressions for the bias and mean squared error (M.S.E.) of $\hat{\xi}_c$ conditional on the event $|\hat{\beta}_1| > 0$. Shukla and Datta (1985) studied the bias and M.S.E. of $\hat{\xi}_c$ under the truncation procedure that $H_0 : \beta_1 = 0$ is rejected with some specified probability α. The argument behind this conditioning is that in practice, it is unlikely that one would proceed to estimate ξ unless one is convinced that $\beta_1 \neq 0$. One way is to test the hypothesis $H_0 : \beta_1 = 0$ and estimate ξ only if H_0 is rejected. This suggests a reasonable truncation procedure of the distribution of $\hat{\beta}_1$ around zero. Perng and Tong's (1974) two-stage sequential approach is in this spirit (see Section 2.7).

The main support of the classical estimator $\hat{\xi}_c$, has been the consistency of its conditional mean within the assumed model (Berkson, 1969) and the fact that if one assumes normal errors, $\hat{\xi}_c$ is the maximum likelihood estimator. The main disadvantage of the classical estimator has been the fact that it has an undefined mean and infinite M.S.E. for fixed $(x_1, x_2, \ldots x_n)$ and finite n (Williams, 1969b). However, Miwa (1985) showed that $\hat{\xi}_c$ has lower moments conditional on $H_0 : \beta_1 = 0$ being rejected. Consider linear functions of Y in the class $\varphi(Y) = k_0 + k_1 Y$. If k_0 and k_1 are chosen to minimise $\sum_{i=1}^{n} E_F(\varphi(Y_i) - x_i)^2$ where the expectation is with respect to Y for fixed value of $\mathbf{x} = (x_1, x_2, \ldots x_n)$ and F is the distribution function of Y for fixed X, the summation is called the total mean squared error or *compound error* of the calibration experiment. Finding constants k_0 and k_1 that minimise $\sum_{i=1}^{n} E_F(\varphi(Y_i) - x_i)^2$ is the problem of compound estimation and the resulting optimal estimator is called a *linear compound estimator*. Lwin and Maritz (1982) established the asymptotic unbiasedness of the classical estimator on the basis of compound estimation without specific distributional assumptions on the errors in the model. They also showed that in the class of consistent estimators linear in Y, $\hat{\xi}_c$ has asymptotically minimum variance.

2.2.2 The inverse estimator

Krutchkoff (1967) derived an estimator known as the *inverse estimator* by considering the regression of X on Y rather than Y on X for a controlled calibration problem. The estimated regression line is given by

$$\hat{X} = \bar{x} + \frac{S_{xy}}{S_{yy}}(Y-\bar{Y})$$

where $S_{yy} = \sum_{i=1}^{n}(Y_i-\bar{Y})^2$. Let the inverse estimator of ξ be denoted as $\hat{\xi}_I$. This formulation of the problem results in

$$\hat{\xi}_I = \bar{x} + \frac{S_{xy}}{S_{yy}}(\bar{Y}'-\bar{Y})$$

where $\bar{Y}'$ is the mean of the m observations taken at the prediction stage of the calibration process. Krutchkoff concluded on the basis of a Monte Carlo investigation that the M.S.E. of $\hat{\xi}_I$ was uniformly less than that of the classical estimator and so the estimator $\hat{\xi}_I$ was preferable. The Monte Carlo work involved 10,000 repetitions and considered both normal and non-normal error distributions with $n \leq 20$. His paper caused considerable controversy when published because his suggestion went against protocols established as far back as Eisenhart (1939), since the n X-values are fixed in a controlled calibration problem. Williams (1969b) criticised $\hat{\xi}_I$ on the grounds that it was derived on the false assumption that the errors ζ_i (i=1,2,...n) are independent of Y_i in the inverse regression of X on Y given by

$$X_i = \gamma + \delta Y_i + \zeta_i$$

However, Lwin and Maritz (1982) argue that their derivation of $\hat{\xi}_I$ does not require this assumption so William's criticism of $\hat{\xi}_I$ is not justified. Perng (1987) derived the inverse estimator by the method of cross-validation (Stone, 1974). The derivation does not require any specific distributional assumptions.

Berkson (1969), Martinelle (1970), Halperin (1970) and Shukla (1972) pointed out that Krutchkoff's conclusion held only in very restrictive circumstances. Krutchkoff, in a subsequent study (1969) compared the classical and inverse approaches and demonstrated that for a sufficiently large number of observations, the classical method produces a lower M.S.E. outside the range of calibration, however the conclusions drawn in his 1967 paper remain unchanged for X values within the calibration range (i.e. $0 \leq X \leq 1$). Apart from Lwin and Maritz, Hoadley (1970) appears to be the only author to give qualified support to the inverse estimator in that he argued that if one assumed a prior for ξ, a non-central Student density centred at $\bar{x}$ (see equation (2.3.1)) and one assumed a

non-informative joint prior for $(\beta_0,\beta_1,\sigma^2)$, the posterior distribution of ξ had mean $\hat{\xi}_I$ and so the inverse estimator was a Bayes estimator in the case of $m=1$ in model (2.1.2). Brown (1982) maintains that these results imply, in sampling theory terms, that the inverse method will do well if the unknown ξ is reasonably central to the set of pre-chosen X-values of a controlled calibration experiment but not perform so well if ξ happened to be outside this pre-chosen range.

Exact expressions for the mean and M.S.E. of $\hat{\xi}_I$ have been derived recently (Shukla and Datta, 1985; Oman 1985). Prior to these results, Williams (1969b) indicated that the mean, variance and M.S.E. of $\hat{\xi}_I$ are finite for $n \geq 4$. However, $\hat{\xi}_I$ is biased even as $n \to \infty$; the bias is affected mainly by the kurtosis of the error distribution of the model and this effect can be reduced by increasing n (Lwin, 1981). The inverse estimator is unbiased only if $\xi = \bar{x} = \sum_i x_i/n$. The inverse estimator is also an inconsistent estimator (Berkson, 1969).

2.2.3 Comparison of $\hat{\xi}_c$ and $\hat{\xi}_I$

In reply to Krutchkoff's 1967 paper, Berkson (1969) showed that in practice when $|\sigma/\beta_1|$ is small, the asymptotic M.S.E. of $\hat{\xi}_c$ is smaller than the M.S.E. of $\hat{\xi}_I$ except when ξ lies very near to $\bar{x}$. Moreover, the inverse method provides an inconsistent estimator whilst the classical method/approach provides a consistent estimator. Martinelle (1970) obtained the expression for relative efficiency for large n and gives results similar to those of Berkson. Halperin (1970) compared the two estimators on the basis of *Pitman's closeness* criterion, consistency and the M.S.E. of the relevant asymptotic distributions. Suppose $\hat{\xi}_1$ and $\hat{\xi}_2$ are two estimators of ξ, then the Pitman's nearness (PN) of $\hat{\xi}_1$ relative to $\hat{\xi}_2$ is given as

$$PN = Pr\{\,|\hat{\xi}_1 - \xi| < |\hat{\xi}_2 - \xi|\,\}$$

He found that for large n, $\hat{\xi}_I$ was superior to $\hat{\xi}_c$ in the sense of *closeness* only in a closed interval round $\bar{x}$ and inferior elsewhere. In practice this interval seems to be very small. He put forward a family of *modified* inverse estimators $\hat{\xi}(r)$ where $r=1$ gives $\hat{\xi}_I$ but preferred the classical estimator to his modified estimators because $\hat{\xi}_c$ gives an exact interval estimate of ξ whereas the modified estimators do not allow an interval estimate. Krutchkoff (1971) carried out further simulations using the Pitman closeness criterion but these did not lead to any clear-cut decisions. Rothman (1968) and Saw (1970) were also very critical of Krutchkoff's 1967 paper but perhaps the most incisive reply came from Williams (1969b). As previously stated, the main objection to using $\hat{\xi}_c$ is that its M.S.E. is infinite for fixed $(x_1,x_2,\ldots x_n)$ and finite n. Williams showed that if σ^2 is assumed known and normal errors are assumed in model (2.1.1) a unique

unbiased estimator of ξ that is a function of the minimal sufficient statistics can be found. Suppose this is denoted by $\hat{\xi}_M$ and suppose further that $\tilde{\xi}$ is any other unbiased estimator of ξ. Then by the Rao-Blackwell theorem

$$Var\ (\hat{\xi}_M) \leq Var\ (\tilde{\xi})$$

Williams showed that $\hat{\xi}_M$ had infinite variance so establishing that no unbiased estimator of ξ would have finite variance. He concluded that the minimisation of M.S.E. as a criterion for choosing between estimators of ξ was not sensible in the calibration situation. His arguments may be taken as a reply to the objection of using the classical estimator.

A second reply to this objection is that one can always apply a truncation procedure to make both the expectation and M.S.E. of $\hat{\xi}_c$ finite. Several authors have suggested that the performance of the classical estimator should be assessed on conditioning on the event $|\hat{\beta}_1| > 0$. Using Tchebycheff's inequality

$$P\left(|\hat{\beta}_1 - \beta_1| \geq k\right) \leq \frac{\sigma^2}{k^2 \beta_1^2 S_{xx}} \qquad k > 0$$

By making S_{xx} large and providing $|\sigma/\beta_1|$ is not large, the probability that $\hat{\beta}_1$ lies in an interval which contains very small values including zero can be made very small. So it is possible, by increasing n and choosing values of X which are not very close to each other, to truncate the distribution of $\hat{\beta}_1$ in such a way as to exclude values of $\hat{\beta}_1$ very close to zero. Krutchkoff used a truncated version of the classical estimator, which replaced $\hat{\beta}_1$ with 0.001 whenever $|\hat{\beta}_1|<0.001$, to obtain the variance and bias of $\hat{\xi}_c$ by means of convergent expansions (1967, 1968). As mentioned in Section 2.2.1, Shukla (1972) and Shukla and Datta (1985) used a *conditional set-up*. Lwin and Maritz (1982) derived a linear compound estimator conditional on the event $|\hat{\beta}_1| > 0$.

When considering M.S.E., the advantage of $\hat{\xi}_I$ over the truncated version of the classical estimator is most pronounced when n is small, $|\sigma/\beta_1|$ is large and when estimation of ξ is restricted to the calibration range i.e. $\min_i X_i \leq \xi \leq \max_i X_i$ (Krutchkoff, 1967, 1968, 1969; Berkson 1969, Martinelle 1970 and Shukla 1972). Lwin (1981) extended Shukla's results (1972) to a more general location and scale family distribution for the errors in model (2.1.1). Lwin gives a detailed comparison of $\hat{\xi}_c$ and $\hat{\xi}_I$ with graphs of their approximate M.S.E.'s as functions of ξ. He showed that the inverse estimator would be worse than the classical estimator as ξ moves further away from $\bar{x}$ for symmetric error distributions and except in the case of extremely peaked error distributions, the inverse estimator would be more efficient at $\xi = \bar{x}$ than the classical estimator (compare Berkson

1969, Halperin 1970). Lwin argued that the M.S.E. criterion could not be dismissed outright without further consideration of the *conditioning* imposed on it.

The conditional or truncated classical estimator derived by Shukla and Datta (1985) compares very favourably with the inverse estimator. Although an earlier conclusion about $\hat{\xi}_I$ and $\hat{\xi}_c$ still holds good for their conditional classical estimator (i.e. $\hat{\xi}_I$ has a smaller M.S.E. than $\hat{\xi}_c$ when ξ is close to the design mean $\bar{x}$) their conditional estimator has a small bias, is consistent and has a finite M.S.E.

Ali and Singh (1981) derived an alternative estimator $\hat{\xi}_a$ which they claimed is uniformly better than either $\hat{\xi}_c$ or $\hat{\xi}_I$. It is given by $\hat{\xi}_a = \lambda\hat{\xi}_c + (1-\lambda)\bar{x}$ where $0 \leq \lambda \leq 1$. Turiel et al. (1982) performed a simulation study in which they compared M.S.E., Pitman's closeness and probability of over-estimation for the classical estimator, the inverse estimator and a modified version of Naszodi's estimator (Naszodi, 1978, see Section 2.7). They considered both small sample and moderately large sample situations.

Chow and Shao (1990) studied the difference between $\hat{\xi}_I$ and $\hat{\xi}_c$, in particular they examined the probability that the ratio of $\hat{\xi}_I$ to $\hat{\xi}_c$ differs from unity by more than a specified small constant. Their results showed that this probability increases as $|\beta_1/\sigma|$ decreases. They proposed methods of estimating this probability.

2.3 THE BAYESIAN APPROACH TO CALIBRATION

Hoadley's (1970) paper was a milestone along the development road of statistical calibration theory because it clearly defined a Bayesian approach to linear calibration. Let $\mathbf{z} = \{(x_i, Y_i) : i=1,2,\ldots n\}$.

Hoadley argued in his paper that the classical estimator $\hat{\xi}_c$ was unsatisfactory from a point of view independent of M.S.E. considerations. The usual F-statistic for testing $H_0 : \beta_1 = 0$ is given by $F = \dfrac{\hat{\beta}_1^2 S_{xx}}{\hat{\sigma}_1^2}$ where

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 + \sum_{j=1}^{m}(Y_j' - \bar{Y}')^2}{(n+m-3)}$$

If F is much larger than $F_{\alpha;1,(n+m-3)}$ (the upper $100(1-\alpha)\%$ point of the F distribution with 1 and $(n+m-3)$ degrees of freedom), then $\hat{\xi}_c$ is fairly precise an estimator but if β_1 is close to zero and H_0 cannot be rejected, then $\hat{\xi}_c$ is very imprecise, i.e. the data of the calibration experiment, $\mathbf{z}$, contain information about the precision of $\hat{\xi}_c$. It would seem reasonable to have some way of giving less weight to $\hat{\xi}_c$ when it is known to be unreliable. This is precisely what a Bayes estimator does. Hoadley argued that conditioning on the data of the calibration

experiment was the right way to approach the calibration problem.

Hoadley took a general form of prior density

$$p(\beta_0,\beta_1,\sigma^2,\xi) \propto p(\beta_0,\beta_1,\sigma^2)p(\xi)$$

and obtained explicit results using the conventional non-informative prior distribution for $(\beta_0,\beta_1,\sigma^2)$ i.e. $p(\beta_0,\beta_1,\sigma^2) \propto \sigma^{-2}$. He assumed normal errors in models (2.1.1) and (2.1.2) and showed that the posterior density of ξ is given by

$$p(\xi|Y',z) \propto p(\xi)L(\xi)$$

where $L(\xi)$ is the predictive density function of $\bar{Y}'$. Hoadley remarked that $L(\xi)$ is a kind of likelihood function representing the information about ξ from all sources except the prior distribution of ξ. However $L(\xi)$ is not integrable so it is necessary that $p(\xi)$ is a proper density function so that overall integrability is achieved, enabling evaluation of a proper posterior density for ξ.

Halperin (1970) recognised the inverse estimator $\hat{\xi}_I$ as a Bayes estimator corresponding to a normal linear model for the calibration experiment and a normal prior density for ξ. Dunsmore (1968) derived $\hat{\xi}_I$ as the Bayes estimator of ξ corresponding to a bivariate normal distribution of (X,Y) and a non-informative prior density function

$$p(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) \propto \sigma_1^{\upsilon-4}\sigma_2^{\upsilon-4}(1-\rho^2)^{\frac{1}{2}\upsilon-3} \qquad \upsilon \leq n$$

for the set of parameters involved in the bivariate normal model. He considered the case of $m=1$ in model (2.1.2), i.e. one observation at the prediction stage. Hoadley also showed that if one assumes a prior for ξ of the form

$$p(\xi) = St\left[n-3,\ \bar{x},\ \left[1+\frac{1}{n}\right]\frac{S_{xx}}{n-3}\right] \qquad (2.3.1)$$

which is the density function of a Student distribution based on $(n-3)$ degrees of freedom, centred at $\bar{x}$ and with scale factor $\left[1+\frac{1}{n}\right]\frac{S_{xx}}{n-3}$, then if $m=1$ in model (2.1.2), the posterior mean is $\hat{\xi}_I$ and the scale factor depends on F defined above. In Bayesian terms, $\hat{\xi}_I$ can be thought of as a shift of $\hat{\xi}_c$, the classical estimator, towards the prior mean, $\bar{x}$. The more informative the data (i.e. the larger F) the smaller the shift. Using Hoadley's relationship,

$$\hat{\xi}_I - \bar{x} = R\left[\hat{\xi}_c - \bar{x}\right]$$

where $R = \{1 + (n-2)/F\}^{-1}$ and F is as defined above, it can be seen that in all situations

$$|\hat{\xi}_c - \bar{x}| \geq |\hat{\xi}_I - \bar{x}|$$

with equality only when $R = 1$ i.e. there is no error in model (2.1.1). So the more informative the data, the more one moves from the prior mean towards the estimate $\hat{\xi}_c$. Since, subject to very weak conditions, Bayes estimators are consistent and the inverse estimator, $\hat{\xi}_I$, is inconsistent (Berkson, 1969), Hill (1981) argued that the inverse estimator was only Bayes for $m = 1$. If one uses the prior for ξ given by equation (2.3.1) and Hoadley's likelihood function $L(\xi)$ for the case of $m > 1$, the resulting Bayes estimator (i.e. the posterior mean/mode) is consistent as n,m both tend to infinity (Brown, 1982). Brown remarks that the posterior mean or mode may be regarded as the correct generalisation of the Krutchkoff (inverse) estimator to $m > 1$. Hill (1981) conjectures that this Bayes estimator will behave asymptotically like the maximum likelihood estimator as m and n go to infinity at appropriate rates. Brown (1982) obtained multivariate extensions of both Hoadley's theorems (see Section 2.6). Here the term multivariate is intended to mean multivariate in both the X's and the Y's.

Aitchison and Dunsmore (1975) assumed normal errors and a non-informative prior for $(\beta_0 + \beta_1\xi, \sigma^2)$. Through a predictive distribution approach, they obtained the posterior distribution of ξ which they call the calibrative distribution. Their results agree with those of Hoadley when $m = 1$ but for $m > 1$ the predictive density function, $L(\xi)$, is a Student-Siegel density function. The form of $L(\xi)$ is as follows

$$L(\xi) \propto \frac{V_2^{\frac{(m-3)}{2}}}{[V_1 c]^{\frac{1}{2}} V_1^{\frac{(m-1)}{2}} \left[1+\frac{(\bar{Y}'-\hat{\beta}_0-\hat{\beta}_1\xi)^2}{V_1 c}+\frac{V_2}{V_1}\right]^{\frac{\nu+1}{2}}} \qquad (2.3.2)$$

where $V_1 = \sum_{i=1}^{n}(Y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2$, $V_2 = \sum_{j=1}^{m}(Y_j' - \bar{Y}')^2$, $c = \left[\frac{1}{m}+\frac{1}{n}+\frac{(\xi-\bar{x})^2}{S_{xx}}\right]$

and $\nu=n+m-3$. This should be compared with $L(\xi)$, using Hoadley's choice of prior for $(\beta_0,\beta_1,\sigma^2)$ which is as follows

$$L(\xi) \propto \frac{1}{[(V_1+V_2)c]^{\frac{1}{2}}\left[1+\frac{(\bar{Y}'-\hat{\beta}_0-\hat{\beta}_1\xi)^2}{(V_1+V_2)c}\right]^{\frac{\nu+1}{2}}} \qquad (2.3.3)$$

In many cases, the calibrative density function is not of standard form because of the prior density $p(\xi)$ and has to be constructed using numerical integration techniques. However, Aitchison and Dunsmore showed that by choosing suitable (tractable) priors for ξ, the calibrative density functions are non-central Student

density functions with the following means

$$\hat{\xi}_I \qquad \text{when } m = 1$$

and $$\hat{\xi}_B = \bar{x} + \frac{S_{xy}}{S_{yy} + \sum_{j=1}^{m}(Y_j' - \bar{Y}')^2}(\bar{Y}' - \bar{Y}) \quad \text{when } m > 1$$

The calibrative density function for the case $m = 1$ is identical to that derived by Hoadley (1970).

Aitchison and Dunsmore compare (for the case of $m > 1$) the Bayes estimator, $\hat{\xi}_B$, the classical estimator, $\hat{\xi}_c$, the inverse estimator, $\hat{\xi}_I$ and Halperin's modified inverse estimator. They note that $\hat{\xi}_B$, unlike the three other estimators, takes account of the variation in the m future observations $Y_1', Y_2', Y_3', \ldots Y_m'$ in such a way that the more variation there is in these observations, the nearer the predictive mean is to the prior mean, $\bar{x}$. Aitchison and Dunsmore (1975) argue that this is a good property for an estimator to have because one should be reluctant to change one's prior views if additional data are very variable.

Hunter and Lamboy (1981a) derived another approach to the calibration problem. For the case of σ^2 known, they assumed normal errors and non-informative priors for (β_0, β_1) and for η where $\eta = \beta_0 + \beta_1 \xi$. In particular, they assumed the priors $p(\beta_0, \beta_{1)}$ and $p(\eta)$ were locally uniform. For the case of σ^2 unknown, they again assumed normal errors and a non-informative prior for $(\beta_0, \beta_1, \sigma^2, \eta)$ of the form,

$$p(\beta_0, \beta_1, \sigma^2, \eta) \propto \sigma^{-2}$$

They further postulated *a priori* independence between (β_0, β_1) and η. The first stage of their analysis was to derive the posterior distribution of (β_0, β_1, η) and then using equation (2.1.3) they derived the posterior distribution of ξ. The posterior densities of ξ are the posterior densities of a ratio of bivariate normal random variables (in the σ^2 known case) and of bivariate t random variables (in the σ^2 unknown case). The posterior distribution of ξ, thus derived, has infinite variance but according to the authors this presents no problems in practice. The authors obtained Bayesian HPD (highest posterior density) intervals for ξ in a particular practical problem and they laid down conditions under which these Bayesian intervals could be approximated by Fieller's (1954) confidence intervals. They suggested using the posterior mode as a point estimate for ξ since they thought it was logical that the point estimate be contained in the reported $100(1-\alpha)\%$ HPD interval(s) and the only estimate for which this criterion is satisfied for all α $(0 \leq \alpha \leq 1)$ is the mode.

The posterior distribution derived by Hunter and Lamboy differs from the Bayesian posterior distribution found by Hoadley (1970) because Hoadley does not treat ξ as an explicit function of η,β_0,β_1 as Hunter and Lamboy do, but directly uses a prior distribution for ξ. In Hunter and Lamboy's analysis, the prior for ξ is implicitly given by the priors for η,β_0,β_1. Another fundamental difference in the two Bayesian approaches is that Hoadley assumes an *a priori* independence between ξ and (β_0,β_1) whereas Hunter and Lamboy assume an *a priori* independence between η and (β_0,β_1). Several authors (Hill, 1981; Lawless, 1981; Lwin, 1981 and Orban, 1981) have argued that Hoadley's independence assumption is more natural and preferable, especially when there is non-negligible prior information about ξ.

Hunter and Lamboy's paper stimulated six discussion papers (Hill, 1981; Lwin, 1981; Rosenblatt and Spiegelman, 1981; Easterling, 1981; Orban, 1981 and Lawless, 1981). Hill (1981) was very critical of Hunter and Lamboy's paper. He criticised, amongst other things, their choice of priors for η and (β_0,β_1), their assumption of *a priori* independence between η and (β_0,β_1) and the lack of information on the location, shape and behaviour of the posterior distribution of ξ. Hill commented that Hunter and Lamboy's analysis is a special case of Hoadley's (1970). He argued that their posterior distribution of ξ is the Hoadley posterior distribution that would arise from using the modified prior density $p(\beta_0,\beta_1,\sigma^2,\xi)$ proportional to $|\beta_1|/\sigma^2$ in the case of σ^2 unknown and $|\beta_1|$ if σ^2 is known. Hill (1981) and Lawless (1981) argued for a family of proper prior distributions which included Hunter and Lamboy's choice of priors as a limiting case. Hill felt that the choice of a family of gamma $(Ga(\alpha,\lambda),\alpha,\lambda>0)$ prior distributions for β_1, with α and λ chosen to reflect prior knowledge, might yield a more satisfactory analysis, including a consistent estimator of ξ with finite M.S.E. Lawless (1981) made the comment that it was not clear how to incorporate prior information about ξ using Hunter and Lamboy's approach. Lwin (1981) argued that any Bayesian approach to calibration should incorporate a *conditioning* of the parameter space in the same way that Shukla (1972) conditioned on the event $|\hat{\beta}_1|>0$. A criticism of Hunter and Lamboy's prior on η which came in a later paper (Brown, 1982) was that there was no natural generalisation of this prior to several future Y' observations corresponding to different unknown ξ's.

In the practical example considered by Hunter and Lamboy, the midpoint of the $100(1-\alpha)\%$ HPD intervals for ξ is the maximum likelihood estimate of ξ. This occurs for $\alpha = 0.1$, 0.05 and 0.01 and was noted by Hill (1981). It led him to conjecture that as n and m go to infinity, Hunter and Lamboy's posterior distribution of ξ tends to concentrate near the maximum likelihood estimate. In

response, Hunter and Lamboy (1981b) stated that they felt it was only necessary for n to go to infinity and that in the limit the posterior distribution would concentrate at (or near) the maximum likelihood estimate.

It will be recalled that $\hat{\xi}_c$, the classical estimator, is the maximum likelihood estimator if one makes the usual assumption of normal errors in models (2.1.1) and (2.1.2). Hill (1981) and Orban (1981) suggested that the Hunter and Lamboy form of prior distribution for η (and implicitly for ξ) tended to provide Bayesian support for the classical estimator, $\hat{\xi}_c$, whereas Hoadley's prior distribution for ξ, outlined above in equation (2.3.1), tended to provide Bayesian support for the inverse estimator, $\hat{\xi}_I$. This suggestion has been confirmed by Brown (1982) in his paper on multivariate calibration. In particular he stated that he saw the classical solution as valid only if the distribution of ξ is thought to be rather flat and wider than the designed distribution of **X** and Hunter and Lamboy's implied prior for ξ is vague and flat.

Hunter and Lamboy (1981b) give a detailed response to all the discussion papers together with a Bayesian interpretation of Fieller's theorem (1954). This paper shows (a) one way in which a Fieller region can be regarded as a Bayesian region and (b) why the proper Bayesian HPD region is preferred and reported to any client. The latter, i.e. (b) is especially important when there are substantial differences between the two results.

In bioassays or enzyme assays, the calibration curves (i.e. concentration response curves) are usually non-linear. Racine-Poon (1988) used a Bayesian approach to a non-linear calibration problem arising from agro-chemical bioassays. If ξ represents the unknown concentration of a new sample presented at the prediction stage, the posterior distribution of ξ can be calculated by using an efficient numerical integration technique such as Naylor and Smith's (1982). Racine-Poon pointed out that in practice, when concentrations of hundreds of samples have to be determined routinely, the numerical effort can be prohibitive and therefore proposed an approximation method to reduce the calculation. The combination of two or more bioassays is also discussed.

Smith and Corbett (1987) applied Bayesian and maximum likelihood methods to the estimation of the length of a marathon course. The data came from a detailed report on the course measurement of the 1984 Olympic Marathon. Apart from considering a multivariate linear regression model, they consider a dynamic model where the calibration constants change with time. Finally, Spezzaferri (1985) used a Bayesian approach to develop a method of choosing among K different multivariate calibration experiments associated with K different instruments.

2.4 MULTIPLE USE OF THE CALIBRATION CURVE

Rosenblatt and Spiegelman (1981) in their discussion on calibration consider it important to distinguish the case where the calibration curve is used only once and the case where it is used repeatedly and interval estimates are reported for a series of determinations. In bioassay a *standard curve* is constructed on which all future assays (calibrations) are to be run.

If we refer back to the chemist's problem in Chapter 1, suppose we are presented not with one new sample at the second stage of the calibration process but with K new samples and suppose we make measurements $Y_1^*, Y_2^*, Y_3^*, \ldots Y_K^*$, one on each sample, and we require interval estimates for the unknown levels of chemical A $(\xi_1, \xi_2, \xi_3, \ldots \xi_K)$ in each of the K samples presented. This problem involves multiple use of the calibration curve which has been derived at the first stage of the calibration problem.

The problem was first treated by Mandel (1958) and another solution was given by Miller (1966). When K is unknown and possibly arbitrarily large, the results of Mandel and Miller do not apply. Lieberman et al. (1967) considered the problem of K unknown or very large. They constructed simultaneous confidence intervals called *unlimited simultaneous discrimination* intervals. (The classical discriminant problem *could* be regarded as a special case of calibration with X taking one of a finite number of values. However, here we are using calibration to mean that X is a continuous variable so it is possibly confusing to use the word *discrimination*). Lieberman et al. (1967) assumed a linear model as given in equation (2.1.1) where the ε_i are independent $N(0,\sigma^2)$. For each set of constructed intervals for $\xi_1, \xi_2, \xi_3, \ldots \xi_K$, there are two probabilities P and α. This is because there are two sources of uncertainty, there is the uncertainty associated with the outcome of the calibration experiment and there is uncertainty that can be attributed to errors in the future measurements $Y_1^*, Y_2^*, \ldots Y_K^*$. Consider for example the case of $P = 0.90$ and $\alpha = 0.05$. Suppose one performs a calibration experiment, estimates a regression line and then constructs K intervals for $\xi_1, \xi_2, \ldots \xi_k$ based on this regression line and measurements $Y_1^*, Y_2^*, \ldots Y_K^*$. The K intervals have the property that at least 90% of the intervals will contain their respective ξ's with confidence 0.95. So if the same calibration procedure is repeated many times, producing different estimated regression lines and each time K intervals for $\xi_1, \xi_2, \ldots \xi_K$ are constructed, then for 95% of the calibration procedures at least 90% of the intervals will contain their correct ξ's. For each of the remaining 5% of calibration procedures, the percentage of intervals enclosing their ξ's may be greater or less than 90%. Lieberman et al. used two methods to obtain simultaneous intervals. The generally more efficient method of the two

was based on application of the Bonferroni inequality to a confidence band for the regression line and a confidence interval for the unknown standard deviation σ.

By way of explanation of Bonferroni inequalities consider a family of statements $\mathcal{F} = \{S_f\}$ where $N(\mathcal{F})$ is the number of statements in the family. The statements might be statements concerning hypotheses or they might be confidence statements. Define the indicator function, $I(S_f)$ as follows;

$$I(S_f) = \begin{cases} 1 & \text{if } S_f \text{ is incorrect} \\ 0 & \text{if } S_f \text{ is correct} \end{cases}$$

Let $\alpha_f = P\{I(S_f) = 1\}$, $f = 1,2,\ldots N(\mathcal{F})$.

Then a Bonferroni inequality is an inequality given by

$$P\{\bigcap_f [I(S_f) = 0]\} \geq 1-\alpha_1-\alpha_2-\ldots-\alpha_{N(\mathcal{F})}$$

Lieberman et al. (1967) used the Bonferroni inequality with $f = 2$. If $f = 2$, the inequality becomes

$$P\{S_1 \text{ is correct and } S_2 \text{ is correct}\} \geq 1 - P(S_1 \text{ is incorrect}) - P(S_2 \text{ is incorrect})$$

In their analysis, S_1 was a statement about σ with confidence $1-\frac{1}{2}\alpha$ and S_2 was a statement about the regression line of Y on X with confidence $1-\frac{1}{2}\alpha$. The Bonferroni inequality (with $f = 2$) was used to combine the two statements so that both held with confidence at least $(1-\alpha)$.

In the same spirit as Lieberman et al. (1967), Clark (1979) used Bonferroni inequalities to obtain simultaneous calibration intervals for $\xi_1,\xi_2,\ldots,\xi_K$ given observations $Y_1^*,Y_2^*,\ldots,Y_K^*$ at the prediction stage and the model

$$Y_j^* = F(\xi_j) + \varepsilon_j' \qquad j = 1,2,\ldots K$$

It was assumed that the ε_j' were independent $N(0,\sigma_j^2)$ $j=1,2,\ldots K$ and F was some smooth but unknown function, estimated by estimators of the form (see Section 2.5)

$$\hat{F}(x) = \mathbf{u}(x)^T\mathbf{Y}$$

Scheffé (1973) greatly extended this approach to calibration via simultaneous intervals. Scheffé assumed that for each i, Y_i is $N(m(X_i,\beta),\sigma^2)$, where $m(X,\beta) = \sum_{j=1}^{p}\beta_j g_j(X)$ for $\min_i X_i \leq X \leq \max_i X_i$. The $\{\beta_j\}$ are unknown parameters and the $\{g_j(X)\}$ are known functions with continuous derivatives. The $n \times p$ matrix whose (i,j)th element is $g_j(X_i)$ is of rank p. So the equation of the estimated calibration curve is given by

$$\hat{Y} = \sum_{j=1}^{p} \hat{\beta}_j g_j(X) = m(X,\hat{\beta}) \qquad (2.4.1)$$

where $\hat{\beta}_j$ are the least squares estimates of β_j $(j=1,2,...p)$. For each Y-value taken at the prediction stage, the Scheffé procedure gives a lower limit ξ_L, a point estimate, and an upper limit ξ_U. Associated with these intervals there are two probabilities α and δ. We require bounds for the calibration curve that will contain the entire curve with a pre-chosen probability $(1-\delta)$ so that δ can be thought of as describing the uncertainty level associated with the outcome of the calibration experiment. The second probability level α can be thought of as describing the uncertainty level attributable to errors in future measurements $Y_1^*, Y_2^*, \ldots, Y_K^*$. By the Scheffé procedure the probability is $\geq 1-\delta$ that at least $(1-\alpha)$ of the K intervals constructed will contain the correct ξ's (i.e. $\xi_1, \xi_2, \xi_3, \ldots, \xi_K$).

Consider constructing an interval for ξ where the corresponding Y-value is Y^*. If σ were known, we could state that Y, corresponding to the unknown ξ lies within the $(1-\alpha)$ confidence interval $(Y^* - z_{1-\frac{1}{2}\alpha}\sigma,\ Y^* + z_{1-\frac{1}{2}\alpha}\sigma)$ where z is $N(0,1)$. The Scheffé procedure expands this interval appropriately to account for the facts that σ is estimated and that this estimate of σ is used for the $(1-\delta)$ bounds on the calibration curve and for bounds on many different future Y-values. It then combines the $(1-\alpha)$ confidence interval for Y with the $(1-\delta)$ bounds on the calibration curve to produce calibration intervals $I(y)$ for each unknown ξ. The lower and upper curves (f_{lower} and f_{upper} respectively) described by Scheffé for monotonic increasing calibration curves have the form

$$f_{\text{lower}} = m(X,\hat{\beta}) + \hat{\sigma}(c_1 + c_2 S(X))$$

$$f_{\text{upper}} = m(X,\hat{\beta}) - \hat{\sigma}(c_1 + c_2 S(X))$$

where $S(X) = [\text{standard deviation of } m(X,\hat{\beta})]/\sigma$. Then,

$$I(Y) = [f_{\text{upper}}^{-1}(Y),\ f_{\text{lower}}^{-1}(Y)].$$

For example if $Y = Y^*$, the calibration interval for ξ is given by

$$f_{\text{upper}}^{-1}(Y^*) \leq \xi \leq f_{\text{lower}}^{-1}(Y^*) \quad \text{i.e.} \quad \xi_L = f_{\text{upper}}^{-1}(Y^*) \quad \text{and} \quad \xi_U = f_{\text{lower}}^{-1}(Y^*)$$

If σ is known then $c_1 = z_{1-\frac{1}{2}\alpha}$ and $c_2 = \chi^2_{1-\delta}(v_1)$ where $\chi^2_{1-\delta}(v_1)$ is the upper δ point of the χ^2-distribution and v_1 is the regression degrees of freedom. If σ is not known, the choice c_1 and c_2 is considerably more difficult. Scheffé gives details of the formulae for c_1 and c_2 when σ is unknown and a formula for calculating $S(X)$. He also explains the use of his procedure in detail for a linear calibration curve. Scheffé compared his intervals with the Bonferroni intervals of

Lieberman et al (1967).

In 1973, Oden also wrote a paper on simultaneous confidence intervals in inverse linear regression. Like Lieberman et al. (1967) and Scheffé (1973) his treatment involved the use of two specified probabilities which he called p and δ. Carroll and Spiegelman (1988) put forward a method based on a modification of the Scheffé (1973) confidence statements which they maintain is easier to implement than Scheffé's and generally leads to shorter intervals. They applied their method to a calibration problem in atomic absorption spectroscopy and compared their results with those obtained using Scheffé's procedure.

Scheffé's approach to calibration via tolerance regions has been criticised by Lindley (1972). Lindley argues that they violate the likelihood principle and are therefore not acceptable. The likelihood principle states that if $\mathbf{x}_1, \mathbf{x}_2$ are two data sets arising from two different experiments and the likelihoods are the same apart from a multiplicative constant i.e.

$$L_{\mathbf{x}_1}(\theta) = C_{\mathbf{x}_1,\mathbf{x}_2} L^{+}_{\mathbf{x}_2}(\theta) \qquad (2.4.2)$$

then data $\mathbf{x}_1$ and $\mathbf{x}_2$ are equivalent for all purposes of inference about θ, the unknown parameter. It should be noted that the constant C does not depend on θ. The principle follows from the Bayesian argument since if equation (2.4.2) holds then

$$p(\theta|\mathbf{x}_1) = p(\theta|\mathbf{x}_2) \quad \text{for all } \theta$$

i.e. the posterior distributions of θ based on data $\mathbf{x}_1, \mathbf{x}_2$ are identical so any inferences about θ will be identical. A result which follows from this principle is that only events that have in fact occurred need to be considered in any inferences and decisions. Hence classical significance tests do not satisfy the likelihood principle because they involve calculating probabilities of what has *not* been observed in addition to what has been observed. Confidence intervals likewise violate the likelihood principle. Tolerance intervals represent an extension of the confidence concept from a statement about a parameter to one about a future observation, or a set of future observations which necessitate reference to some property of χ, the sample space, other than the observed $\mathbf{x}$ ($\mathbf{x} \in \chi$) and so violate the likelihood principle. Lindley (1972) commented that such an approach produces tortuous statements which are difficult to comprehend and are replaced in the Bayesian analysis, by a single statement that is much simpler. The Bayesian approach involves predictive distributions.

Brown (1982) is in accord with Lindley's views. In response to the discussion of his paper on multivariate calibration, he commented that an interval for an unknown ξ corresponding to an observed $\mathbf{Y}'$ at the second stage of the

calibration process should not depend on other as yet unknown and unobserved $(\xi, \mathbf{Y}')$. A set of observed $\mathbf{Y}'$ at different unknown ξ provide information on the distribution of future ξ. As soon as two or more $\mathbf{Y}'$ are observed corresponding to different but unknown ξ, it is possible to update the posterior distribution of ξ (as defined by Hoadley (1970) in the univariate case) by using Bayes theorem. The updating involves calculating posterior predictive distributions. Obviously if one rejects Scheffé's approach to calibration via simultaneous calibration intervals for the unknown ξ's, one must update the calibration (either continuously or at regular intervals as is practically expedient) as Y' are observed corresponding to different ξ's. Copas (1982) put forward a method for obtaining an updated point estimate of ξ, in random calibration, given a set of observed Y' corresponding to different ξ's. However, it should be pointed out that sometimes updating is not always possible.

2.5. NON-PARAMETRIC APPROACHES TO CALIBRATION

In recent years at least seven papers have taken a non-parametric approach to calibration. Three of these papers involve T. Lwin, namely Lwin and Maritz (1980), Lwin (1981), Lwin and Maritz (1982). Lwin and Maritz (1980) consider a random calibration experiment in which only the bivariate random variables (X,Y) can be observed (compare Tallis 1969). They derived the distribution of X given Y basing estimation of the marginal distribution of X on the sample distribution function. For point estimation, if $f(y_i|x_i,\beta)$ is the probability density function of Y_i conditional on $X_i = x_i\,(i=1,2,\ldots n)$ then the predictor of X when $Y = Y'$ is observed (at the prediction stage) is $\sum w_i x_i$ where

$$w_i = f(Y'|x_i,\beta) \Big/ \sum_{j=1}^{n} f(Y'|x_j,\beta)$$

Their *non-linear* predictor as they call it is a special member of the following class of predictors:

$$\hat{\xi}(Y') = \sum_{i=1}^{n} x_i\, w_i(Y', \mathbf{z})$$

where $w_i(Y', \mathbf{z})$ are *weights* attached to the x_i's. This class has been studied in detail in terms of consistency properties by Stone (1977). Comparing this with the inverse estimator, $\hat{\xi}_I$, the latter is a predictor linear in Y' whereas Lwin and Maritz do not restrict themselves to the class of predictors linear in the observation Y'. It has the advantage that it is applicable for non-normal error distributions; the error distribution need only be a member of the location and scale family. It is also very flexible as the calibration curve $m(X,\theta)$ can have

many forms, not just linear in X or linear in θ.

Lwin and Maritz (1982) compared the classical and inverse estimators $\hat{\xi}_c$ and $\hat{\xi}_I$ respectively by using a compound estimation approach (see Section 2.2.1). With this approach, $\hat{\xi}_c$, is a linear compound estimator satisfying the criterion of asymptotic unbiasedness whilst $\hat{\xi}_I$ is a linear compound estimator without the unbiasedness constraint. Their formulation required no specific distributional assumptions. Their approach showed that $\hat{\xi}_I$ was superior to $\hat{\xi}_c$ only if the current ξ value was sampled from the same population as previous X values i.e. $x_1, x_2, \ldots x_n$ of the calibration experiment. This paper should be compared with Brown (1979) who obtained an optimal linear predictor $\lambda_0 + \lambda Y$ where the λ_0 and λ were chosen to minimise the integrated mean square error rather than the compound error as with Lwin and Maritz (1982). The integrated mean square error is defined over a range (L,U) with respect to a weight function $w(x)$ as

$$IMSE = \int_U^L MSE(x)w(x)dx.$$

Clark has produced two papers on calibration with particular reference to radio-carbon dating (1979, 1980). His 1980 paper assumes the model

$$Y_{ij} = F(x_i) + e_{ij} \quad i=1,2,\ldots n, \quad j=1,2,\ldots m_i, \quad \Sigma m_i = N$$

where $\{x_i\}$ are known constants, F is some unknown but *smooth* function and $\{e_{ij}\}$ are uncorrelated random variables with zero mean but constant variance σ^2. The $\{x_i\}$ are assumed to be distinct with $x_1 < x_2, \ldots < x_n$. The estimate of F was chosen by cross-validation from a class of estimators defined by

$$\hat{F}(x) = \mathbf{u}(x)^{\mathrm{T}}\mathbf{Y} \tag{2.5.1}$$

where $\mathbf{u}(x) = (u_1(x), u_2(x), \ldots, u_n(x))^{\mathrm{T}}$ is a $n \times 1$ vector of known functions, possibly indexed by some index parameter β, $Y_{i\cdot} = \sum_j y_{ij}/m_i$ $(i=1,2,\ldots n)$ and $\mathbf{Y} = (Y_{1\cdot}, Y_{2\cdot}, Y_{3\cdot}, \ldots, Y_{n\cdot})^{\mathrm{T}}$. This class of estimators includes as special cases least-squares polynomials and splines, approximating and interpolating splines and the CS-estimators. The CS-estimators are of particular interest to Clark and the generalised CS-estimator is defined as

$$\hat{F}(x) = \int_{-\infty}^{\infty} h(t) \frac{1}{b(x)} W\left[\frac{x-t}{b(x)}\right] dt$$

where h is the unique first-degree natural spline interpolating data $\{(x_i, Y_{i\cdot})\}$, W is an arbitrary density function and b, the bandwidth, is a suitable bounded non-negative function. The cross-validation algorithm used is defined as follows; let $P_1, P_2, \ldots P_k$ be a partition of the index set $U = \{1,2,\ldots n\}$ and let n_j be the number of elements in P_j. Corresponding to each subset P_j , one conceptually

sub-divides the complete sample into an *estimation sample* $\{(x_i,Y_i.)\ \ i \in \bar{P}_j\}$ and a *validation sample* $\{(x_i,Y_i.)\ \ i \in P_j\}$ and for various trial values of β computes

$$C_j(\beta) = \frac{1}{n_j} \sum_{i \in P_j} w_i \{Y_i. - f^*(x_i|j,\beta)\}^2$$

where $f^*(\cdot|j,\beta)$ denotes the estimator defined by β computed from the jth estimation sample *only*, and $w_1, w_2, \ldots w_n$ are suitable non-negative weights. The cross-validation mean-square error (CVMSE) is defined as

$$\bar{C}(\beta) = \frac{1}{k} \sum_j C_j(\beta)$$

The "best" choice of β is then taken to be that minimising $\bar{C}(\beta)$. This CVMSE is also a device for estimating the average bias of the chosen estimator and this is discussed in detail by Clark (1980). Whereas Clark considered the construction of simultaneous calibration intervals in his 1979 paper under the assumption that the estimator $\hat{F}$ (as defined by equation 2.5.1) has negligible bias, in his 1980 paper he considered the *equivalent* problem of construction of prediction intervals (i.e. intervals for Y given $X = x$) under the assumption that $\hat{F}$ has a non-negligible bias. The CVMSE was used to estimate this bias and an adjustment proposed to both prediction and confidence intervals to compensate, to some extent, for the bias in $\hat{F}$.

Lechner et al. (1982) combined Scheffé's (1973) calibration approach with linear splines to produce simultaneous calibration intervals for the liquid volume (v) in large nuclear material processing tanks, given the differential pressure (p) . The authors assumed the upper part of the tank to be composed of $(k+1)$ distinct and known regions. By considering the cross-sectional area, A(x), at a given height x, and its relationship to p and v, the authors concluded that within each of the $(k+1)$ regions, it was reasonable to assume a linear relationship between p and v. The pressure-volume relationship could therefore be modelled as a linear spline with k knots (see Section 3.1.1, Chapter 3). The location of the knots between line segments were determined empirically and confirmed by engineering analysis. Since recalibration would not be feasible after the tank was in use at a nuclear materials processing plant, the Scheffé procedure was used for describing the uncertainty of volume measurements. The authors chose to use a B-spline basis made up of linear B-splines (De Boor, 1978). If $\{v_i\}$ $i=1,\ldots k$ are the interior knot locations, then the B-spline basis consists of the functions $\beta_i(v)$ where $\beta_i(v)$ is the first-degree (piecewise linear) natural spline interpolating the points $\{(v_j,\delta_{ij})\ j=1,2,\ldots k\}$ where $\delta_{ij}=1$ if $i=j$, or 0 if $i \neq j$. The authors applied their theoretical results to 172 observations from calibration runs on a processing tank

and compared their results with those obtained using propagation-of-error techniques (Naszodi 1978).

Knafl et al. (1984) also considered pressure-volume calibration of a nuclear tank and adopted Scheffé's procedure (1973), i.e. resulting calibration intervals had two associated probabilities α and δ. Their non-parametric approach is more general than Lechner et al. (1982). The latter commented that the greater generality of Knafl et al. comes at a cost, namely wider intervals for volume measurements. Since Scheffé's simultaneous intervals tend to be somewhat conservative, it is questionable whether it is worthwhile adopting a more general model.

They proposed a model which states that at each point $x \in [S_0, S_1]$ and for any real t, m_1, m_2 $(m_1 < m_2)$ and real f, the unknown calibration curve f can be written as $f(t) = f(x) + r(t,x)$ with

$$m_1(t-x) \le r(t,x) \le m_2(t-x) \qquad \text{if } t > x$$

$$m_2(t-x) \le r(t,x) \le m_1(t-x) \qquad \text{if } t < x$$

where m_1 and m_2 are specified. If $m_1 > 0$ this model implies that f is increasing at a rate that is bounded above and below. This does not require differentiability of f and thus reflects the abrupt changes in the tank. Let

$$\eta_i = Y_i - \left[\frac{m_1+m_2}{2}\right](x_i - x) \qquad i=1,2,\ldots. n$$

$$\rho(x_i,x) = r(x_i,x) - \left[\frac{m_1+m_2}{2}\right](x_i - x) \qquad i=1,2,\ldots. n$$

The model for the calibration experiment then becomes

$$\eta_i = f(x) + \rho(x_i,x) + \sigma\varepsilon_i \qquad i=1,2,\ldots n.$$

Knafl et al. sought a linear estimate $\Sigma c_i(x)\eta_i$ of $f(x)$. The c_i were chosen to minimise the maximum M.S.E. subject to $\Sigma c_i(x) = 1$, which ensures that the bias term of the M.S.E. is bounded. It was noted that the estimate of $f(x)$ i.e. $\Sigma \hat{c}_i(x)\eta_i$ depends only on averages of η's at common values of the x_i's. Thus replicated data (repeated x_i's) provided an obvious estimate $\hat{\sigma}^2$ of σ^2 based on the replicated data only. It was further assumed that $\varepsilon_i \sim N(0,1)$ thus $\hat{f}(x)$, conditioned on $\hat{\sigma}$, is normal with mean $f(x) + \Sigma \hat{c}_i(x)\rho(x_i, x)$ and variance $\hat{\sigma}^2 \Sigma \hat{c}_i^2(x)$. The upper and lower calibration curves of Scheffé's method are given by

$$\hat{U}(x) = \hat{f}(x) + B(x) + \hat{a}D(x) + \hat{q}\hat{\sigma}$$

$$\hat{L}(x) = \hat{f}(x) - B(x) - \hat{a}D(x) - \hat{q}\hat{\sigma}.$$

where $B(x)$ is the maximum bias of $\hat{f}(x)$ and $D(x)$ is the standard deviation of $\hat{f}(x)$. Although the calculation of $\hat{f}(x)$ and $\hat{c}(x)$ is an easy process on a computer, the calculation of $\hat{a}$ and $\hat{q}$ is long and complicated involving evaluation of double integrals and complicated equations involving integrals. The authors gave a detailed seven-point procedure for the derivation of the calibration intervals as well as an extension of the method to suit *smoother fs*. If Y^* is a new observation corresponding to an unknown ξ, the calibration interval for ξ is given by

$$\hat{U}^{-1}(Y^*) \leq \xi \leq \hat{L}^{-1}(Y^*)$$

2.6 MULTIVARIATE CALIBRATION

Naes and Martens (1984) state that multivariate calibration is a young discipline and Brown's (1982) paper contributed a great deal to this discipline in its early stages. Brown's (1982) paper is an important paper on multivariate calibration not only because of its contents but because it stimulated great interest in this area of calibration.

Brown (1982) adopted both a *classical* approach akin to that of Eisenhart (1939), Williams (1959) and Fieller (1954) and a *Bayesian approach* akin to that of Hoadley (1970). He assumed for both the calibration experiment and the prediction experiment (i.e. the second stage of the calibration process) multivariate linear models. Suppose there are n observations in the calibration experiment, q response variables $Y_1, Y_2, \ldots Y_q$ and p explanatory variables $X_1, X_2, \ldots X_p$ and $q \geq p$.

$$Y = \mathbf{1}\alpha^{\mathrm{T}} + XB + E \qquad (2.6.1)$$

where $Y(n \times q)$ and $E(n \times q)$ are random matrices, X is a $n \times p$ matrix of fixed constants and $\mathbf{1}$ is a $(n \times 1)$ vector of units. B is a $p \times q$ matrix of unknown parameters, α is a $q \times 1$ vector of unknown parameters. The model for the prediction experiment is given by

$$Y' = \mathbf{1}\alpha^{\mathrm{T}} + \mathbf{1}\xi^{\mathrm{T}}B + E' \qquad (2.6.2)$$

where $Y'(m \times q)$, $E'(m \times q)$ are random matrices, ξ is a $p \times 1$ vector of unknowns and $\mathbf{1}$ is a $(m \times 1)$ vector of units. It is wished to draw inferences about ξ. As defined the calibration experiment is a controlled calibration experiment. If $\mathbf{e}_i$ is the ith row of E, it is assumed that $E(\mathbf{e}_i) = 0$, $E(\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}) = \Gamma$ and $\mathbf{e}_i \sim N(\mathbf{0}, \Gamma)$ for $i = 1, 2, \ldots n$. If $\mathbf{e}_j'$ is the jth row of E', the $\mathbf{e}_j'$ satisfy the above also and it is assumed that they are independent of the $\mathbf{e}_i$. Brown pointed out that $\mathbf{X}$ might consist of p variables derived from a smaller set as in polynomial regression. In

this case, ξ is a vector function of the same reduced number of unknowns.

Let S be the $(q \times q)$ residual sum of products matrix, pooled from the calibration and prediction experiments when $m > 1$. Further let

$$\sigma^2(\xi) = \frac{1}{m} + \frac{1}{n} + \xi^T (X^T X)^{-1} \xi$$

Brown showed by using *classical* arguments that a $100(1-\gamma)\%$ confidence region for ξ is all ξ such that

$$(\bar{Y}' - \hat{\alpha} - \hat{B}^T \xi)^T S^{-1} (\bar{Y}' - \hat{\alpha} - \hat{B}^T \xi) / \sigma^2(\xi) \leq \frac{q}{\nu} F_{\gamma,q,\nu} \qquad (2.6.3)$$

where $F_{\gamma,q,\nu}$ is the upper $100(1-\gamma)\%$ point of the standard F distribution on q and ν degrees of freedom ($\nu = n-p+m-q-1$). This is the multivariate form of the confidence region obtained by Fieller (1954) and others who considered $p=q=1$. In standard multivariate linear regression the region corresponds to fiducial limits of Williams (1959). When $p = q$ the above inequality can be written as

$$\| \xi - C^{-1} D \|_C^2 - \text{constant} \leq 0 \qquad \text{where}$$

$$C = \hat{B} S^{-1} \hat{B}^T - k(X^T X)^{-1} = \hat{B} S^{-1} \hat{B}^T - \left[\frac{q}{\nu}\right] F_{\gamma,q,\nu} (X^T X)^{-1}$$

and C is assumed positive definite. So the confidence region defined by inequality (2.6.3) is a closed ellipsoid. As stated in Section 2.2.1, in the univariate case (i.e. $p=q=1$) the confidence region sometimes degenerates into two disjoint semi-infinite lines or the whole real line. This occurs if

$$\frac{\hat{\beta}_1^2 S_{xx}}{\hat{\sigma}^2} < F_{\gamma,1,n+m-3}$$

The condition that C is positive definite is the direct multivariate extension of the univariate condition. If one lets $\gamma \to 1$ in inequality (2.6.3), then $k \to 0$ and this condition on C will be satisfied if $B \neq 0$. The resulting confidence region degenerates to the point

$$\hat{\xi} = (\hat{B} S^{-1} \hat{B}^T)^{-1} \hat{B} S^{-1} (\bar{Y}' - \hat{\alpha}) \qquad (2.6.4)$$

If $p=q=1$, this gives the classical estimator $\hat{\xi}_c$ defined in Section 2.2.1. As equation (2.6.4) arises when considering the regression of **Y** on **X**, $\hat{\xi}$ is usually called the classical estimator of ξ. Brown and Sundberg (1987) have shown that for $q = p$, the estimator $\hat{\xi}$ defined by equation (2.6.4) is the maximum likelihood estimator of ξ. When $q > p$, however, it is not the maximum likelihood estimator, the latter being shifted from $\hat{\xi}$ by an amount depending on the

inconsistency diagnostic

$$R = (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\hat{\xi})^{\mathrm{T}} S^{-1} (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\hat{\xi}) \qquad (2.6.5)$$

The inconsistency diagnostic R measures the relative lack of consistency of a q-variate observation $\mathbf{Y}'$.

Lieftinck-Koeijers (1988) considered the classical estimator given by equation (2.6.4) in the case of p=1. He showed that this estimator has a finite mean if q>2 and a finite M.S.E. if q>4. He gave exact expressions for the mean and M.S.E. in terms of expectations of Poisson variables. Nishii and Krishnaiah (1988) also considered the classical estimator and showed that when Γ is unknown , the mean of $\hat{\xi}$ is finite iff $q \geq p+1$ and the M.S.E. is finite iff $q \geq p+2$. They give the exact moments of $\hat{\xi}$ using expectations of Poisson variables and compare their results with those of Lieftinck-Koeijers.

Oman (1988) considered the case of where ξ in model (2.6.2) is given by $\xi^{\mathrm{T}} = [h_1(\zeta)\, h_2(\zeta) \ldots\ldots h_p(\zeta)]$ for *known* functions h_j (e.g. squares or logarithms of components of ζ) and one wishes to construct a confidence region for the *unknown* ζ corresponding to a future $\mathbf{Y}'$. One approach to the problem is as follows; obtain a $100(1-\gamma)\%$ confidence region for ξ by using expression (2.6.3). Suppose this region is called R. The confidence region C for ζ is then given by $C = \{\, \zeta : \xi(\zeta) \in R \,\}$. Oman points out that a disadvantage of this approach is that although the region R might be nicely behaved, the region C need not be. He put forward an alternative confidence region for ζ and compared his method with asymptotic results obtained by Fujikoshi and Nishii (1984) and Brown and Sundberg (1987). An application to the estimation of gestational age using ultrasound foetal bone measurements is given in the paper.

Brown's Bayesian approach was similar to that of Hoadley. Hoadley (1970) assumed a general form of prior density $p(\beta_0,\beta_1,\sigma^2,\xi) \propto p(\beta_0,\beta_1,\sigma^2)p(\xi)$. Brown assumed

$$\pi(B,\alpha,\Gamma,\xi) = \pi(B,\alpha,\Gamma)\pi(\xi)$$

Brown also assumed $\pi(\xi|X) = \pi(\xi)$ i.e. the controlled X values provide no information on ξ and he assumed a Jeffrey's invariant prior

$$\pi(B,\alpha,\Gamma) \propto |\Gamma|^{-(q+1)}$$

which is the multivariate analogue of Hoadley's non-informative prior for $(\beta_0,\beta_1,\sigma^2)$ i.e. $p(\beta_0,\beta_1,\sigma^2) \propto \sigma^{-2}$. Brown obtained multivariate extensions to both of Hoadley's theorems which provide extra insight into Hoadley's results.

Brown derived a multivariate generalisation ($\hat{\hat{\xi}}$) of the Krutchkoff inverse estimator by regressing $\mathbf{X}$ on $\mathbf{Y}$, which he compared with the classical estimator $\hat{\xi}$

given by equation (2.6.4). Brown's multivariate methods, including one which is a multivariate extension of Lwin and Maritz (1980) are compared on data from a random calibration experiment and data from a controlled calibration experiment. Some important conclusions are drawn. In particular, Brown suggested that it was a good idea to consider each of the p characteristics, $X_1, X_2, X_3, \ldots X_p$, one at time, forgetting the existence of the other $(p-1)$ characteristics. Sundberg (1985) justifies this in some circumstances. Brown's paper is a discussion paper and there is a considerable amount of discussion of his results with further references.

A recent paper by Brown and Sundberg (1987) has considered an approach to multivariate calibration which involves the profile or maximum relative likelihood (Kalbfleisch and Sprott, 1970). Suppose that the model for the calibration experiment is given by model (2.6.1) i.e.

$$Y = \mathbf{1}\alpha^{\mathrm{T}} + XB + E$$

If $\mathbf{e}_i$ is the ith row of E, it is assumed that $E(\mathbf{e}_i) = 0$, $E(\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}) = \Gamma$ and $\mathbf{e}_i \sim N(\mathbf{0},\Gamma)$ for $i=1,2,\ldots n$. This profile likelihood approach entails forming the maximised likelihood "as if ξ were known" and normalising by the likelihood maximised over all unknown parameters α,B,Γ,ξ. Suppose the model for the prediction stage is given by model (2.6.2) with $m=1$. This approach is offered as an alternative to the Bayesian approach (Brown, 1982), avoiding the need for specification of a prior distribution. The profile likelihood is a function of ξ which has a maximum value of one at a maximum likelihood estimate of ξ and the profile log-likelihood is given (up to an additive constant) by

$$-\frac{1}{2}(n+1)\ln(|\hat{\Gamma}(\xi)|) \qquad (2.6.6)$$

$$\text{where} \quad |(n+1)\hat{\Gamma}(\xi)| \propto \frac{n_0(n+1)}{n} + \frac{[R + (\xi-\hat{\xi})^{\mathrm{T}}\hat{B}S^{-1}\hat{B}^{\mathrm{T}}(\xi-\hat{\xi})]}{r(\xi)}$$

$$\text{with} \quad r(\xi) = 1 + \frac{n}{(n+1)}\,\xi^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\xi$$

and R the inconsistency diagnostic given in equation (2.6.5), $\hat{B}$ the least squares estimator of B and S the residual sum of products matrix from the calibration experiment.

The profile likelihood for $m \geq 1$ is given by

$$C\left[\frac{\sigma^2(\xi)}{\sigma^2(\xi) + (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)\,S^{-1}\,(\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)}\right]^{\frac{1}{2}(n+1)} \qquad (2.6.7)$$

where $\sigma^2(\xi) = \frac{1}{m} + \frac{1}{n} + \xi^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\xi$ and S is the residual sum of products matrix, pooled from the calibration experiment and the prediction stage. The

normalising constant is such that C^{-1} is the value of expression (2.6.7) at its maximum. The Bayes posterior density, $\pi(\xi \mid \mathbf{Y}', \mathbf{z})$, obtained when using a vague or non-informative prior for ξ is given by

$$\pi(\xi \mid \mathbf{Y}', \mathbf{z}) \propto \frac{[\sigma^2(\xi)]^{\frac{1}{2}\nu}}{[\sigma^2(\xi) + (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)\, S^{-1}\, (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)]^{\frac{1}{2}(\nu+q)}}$$

where $\nu = n-p+m-q-1$. If one compares the profile likelihood and the posterior density, one can see that they are closely related. The powers of numerator and denominator of Bayes posterior are $\frac{1}{2}\nu$ and $\frac{1}{2}(\nu+q)$ compared with $\frac{1}{2}(n+1)$ and $\frac{1}{2}(n+1)$ in the profile likelihood. Brown and Sundberg obtained likelihood-based confidence regions for ξ which have the intuitively desirable property of expansion with increasing values of R. This should be contrasted with the unnatural behaviour of the classical confidence regions (see inequality 2.6.3) when $q > p$, which expand for decreasing R and get narrower with increasing R sometimes shrinking to a point (see Oman and Wax, 1984, for a practical example of this).

Brown and Sundberg (1989) examine the case of there being more variables than observations ($n < p+q+1$) when assuming a standard multivariate linear regression model given by equation (2.6.1) with no derived variables. They considered the case of $\mathbf{X}_i$ being regarded as fixed (controlled calibration) and $\mathbf{X}_i$ being regarded as random (random calibration). They showed that if $n > q$, the generalised least squares estimator of ξ and the estimated best linear predictor ($\hat{\hat{\xi}}$) are both unique. By way of an example, they used the NIR (near infra-red) data of Fearn (1983).

There are a wide variety of approaches to multivariate calibration. Naes et al. (1986) compare the multiple linear regression (MLR), ridge regression (RR), principal component regression (PCR) and partial least squares regression (PLSR) approaches with particular reference to the calibration of near infra-red (NIR) instruments. In a NIR calibration problem, the instrument response $\mathbf{Y} = (Y_1\, Y_2\, ...\, Y_q)^{\mathrm{T}}$ is called the spectrum. In the NIR problem considered by Fearn (1983), $q = 6$ and $Y_1, Y_2, ... Y_6$ are measurements of the reflectance of NIR radiation at six different wavelengths. There is only one dependent variable X which is the protein content of wheat samples. The NIR data matrix is highly multicollinear and the data set have been used by several authors since 1983 (Farebrother, 1984; Hoerl et al. 1985; Naes et al. 1986 and Brown and Sundberg, 1989). With reference to the PCR and PLSR methods, both methods project the NIR spectrum into a space determined by vectors and use coordinates in this space as regressors with X as regressand. However, while PCR projects into a space estimated by the spectral variables $Y_1, Y_2, ... Y_q$, PLSR projects into a

space determined by both spectrum and X. According to Naes and Martens (1984), the PCR and PLSR methods have given good results, whilst Naes et al. (1986) have commented that PLSR is a serious competitor to PCR and RR for multicollinear data.

Ridge regression is particularly recommended for cases in which the explanatory variables have high intercorrelations. When these high correlations exist, least squares is known to give estimated coefficients which tend to be too large and often cancel out when taken in combination. Hoerl and Kennard (1970) suggested a modification of the usual least squares estimator from $(X^TX)^{-1}X^TY$ to $(X^TX + cI)^{-1}X^TY$ where c is a scalar added to each diagonal element of the explanatory variable correlation matrix before inversion. The resulting ridge estimator, which is biased, can be shown to have better mean square error properties than least squares if c is correctly chosen. Hoerl and Kennard suggested that c be chosen by examining a ridge trace, which is a plot of coefficients against c and looking for the point at which stability is reached.

Fearn (1983) compared the MLR and RR approaches. The paper provoked quite a bit of discussion about whether prediction methods such as RR should be used in the calibration of NIR instruments (Farebrother 1984 and Hoerl et al. 1985). Naes et al. (1986) performed computations on Fearn's data set and found that when the ratio of the number of calibration samples and the number of wavelengths in the NIR spectrum is low i.e. n/q is small, the RR, PLSR and PCR methods which are biased regression methods, gave much better results than MLR.

A paper by Naes and Martens (1984) compared the PCR approach, the PLSR approach and an approach developed from the use of a multivariate model based on Beer's law in chemistry (the field of spectroscopy). Beer's law states that the spectrum, Y, is a linear function of the concentrations $(X_1, X_2, X_3, \ldots, X_p)$ of one or more chemical constituents (p constituents) of the mixture plus measurement noise. Naes and Martens pointed out that the PCR and PLSR methods are random calibration methods since they are both based on regressing the chemical concentrations $(X_1, X_2, \ldots, X_p)$ on linear combinations of Y whereas the Beer's model approach is able to handle both random and controlled calibration situations. On the other hand, it is noted by Naes and Martens that the PCR and PLSR approaches are probably better suited to handle the problem of measurement noise in both X and Y in the calibration experimental data, $\{(\mathbf{X}_i, \mathbf{Y}_i)\ i=1,2,\ldots n\}$.

Naes (1985a) compared the generalised least squares estimator of ξ , $\hat{\xi}_{GLS}$ and the best linear predictor of ξ , $\hat{\xi}_{BLP}$ which arise from the controlled and

random calibration situations respectively. In this paper he assumed that Beer's model held. If the parameters of the model in both the controlled and random situations are replaced by estimates, let $\tilde{\xi}_{GLS}$ and $\tilde{\xi}_{BLP}$ be the resulting estimators. If the estimates of the parameters of the model are maximum likelihood estimates, $\tilde{\xi}_{GLS}$ and $\tilde{\xi}_{BLP}$ coincide with the estimators $\hat{\xi}$ and $\hat{\hat{\xi}}$ in Brown (1982). Naes showed that, if the estimates of the parameters of the model were consistent , then $\tilde{\xi}_{GLS}$ and $\tilde{\xi}_{BLP}$ behave more and more like $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ respectively as $n \to \infty$. He showed that the set of ξ's where $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ have the same M.S.E. is an ellipsoid given by

$$F = \{ \xi : E[(\hat{\xi}_{BLP} - \xi)^T (\hat{\xi}_{BLP} - \xi) \mid \xi] = \mathrm{tr}(B^T \Gamma^{-1} B)^{-1} \}$$

using the notation of model (2.6.1). If F' denotes F and the region inside F, then if $\xi \notin F'$, $\hat{\xi}_{GLS}$ is best otherwise $\hat{\xi}_{BLP}$ is best. Finally he suggested a new predictor which is a combination of $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ involving indicator functions.

The problem of estimating $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ in the best way is treated in Naes (1985b, 1986). Methods based on different assumptions on Σ, the error covariance matrix of the multivariate linear model, are proposed and analysed in these two papers. It is known from practical experiments (e.g.NIR measurements) that the covariance matrix of multivariate measurements from spectrophotometers is often highly multicollinear. This means that the number of variation sources for the spectrum is smaller than the number of variables in the spectrum. It is therefore assumed in Naes (1985b) that the measurement errors have a linear factor structure, i.e.

$$\mathbf{e}_i = P\mathbf{t}_i + \mathbf{e}_i' \qquad i=1,2,\ldots n.$$

where P is a fixed matrix and $\mathbf{e}_i, \mathbf{e}_i'$ and $\mathbf{t}_i$ are random vectors. These "rank reduction" models have proved to be very suitable for multicollinear data and have given good results in NIR spectroscopy (see Naes, 1985b). The model (2.6.1) becomes

$$Y = \mathbf{1}\alpha^T + XB + TP + E'$$

where X, Y, T and E' are random matrices, B and P are fixed matrices (the rows of P may be regarded as spectra of unknown constituents, while B contains spectra of the known constituents). As this is a random calibration problem, Naes obtained expressions for the best linear predictor of ξ i.e. $\hat{\xi}_{BLP}$ and the estimated best linear predictor, $\tilde{\xi}_{BLP}$. A principal component analysis is used to estimate Σ. Naes examined the properties of $\tilde{\xi}_{BLP}$ and discussed the links between

(a) $\tilde{\xi}_{BLP}$ and the predictor based on multiple linear regression of **X** and **Y**

(b) $\tilde{\xi}_{BLP}$ and the PCR predictor.

Naes pointed out that his calibration procedure using $\tilde{\xi}_{BLP}$ allows calibration of two or more constituents in a mixture simultaneously whereas when random calibration is approached by MLR, stepwise MLR or PCR, each constituent must be calibrated separately, i.e. calibration of different constituents in a mixture requires completely separate computations.

2.7. OTHER APPROACHES TO CALIBRATION

Two papers, Kalotay (1971) and Minder and Whitney (1975) approach calibration using structural inference. Kalotay used arguments from structural inference to justify the construction of a likelihood for ξ, essentially by assuming a non-informative prior likelihood for the nuisance parameters β_0, β_1 and σ^2 and integrating over this prior. He compared his results with those of Hoadley (1970). The posterior distribution for ξ derived by Hunter and Lamboy (1981a) is mathematically identical to that of Kalotay (1971). Minder and Whitney (1975) derived a *marginal likelihood* function (MLF) for ξ and then used various approximations to the RMLF (relative marginal likelihood function obtained by normalising the MLF to have a maximum value of one at the maximum likelihood estimate of ξ) to derive confidence regions for ξ. There are mathematical similarities between their MLF for ξ and the profile likelihood function for ξ (see Kalbfleisch and Sprott, 1970, for the latter) but Minder and Whitney maintained that despite these similarities, the two functions could lead to considerably different inferences for small $(n + m)$. Minder and Whitney comment that their MLF for ξ is rather similar to the predictive density function, $L(\xi)$, of Hoadley (1970) and the approximate structural density function of Kalotay (1971).

Tallis (1969) considered the problem of calibration with supplementary information. In addition to the information of a controlled calibration experiment which gives knowledge of the conditional distribution of Y given $X = x$ for selected values of x, Tallis made extra observations on the marginal distribution of Y alone. The combined information is then used to estimate the conditional distribution of X given Y. The attainment of a viable solution depends on finding a unique solution to a Fredholm integral equation of the first kind.

The approach of Muhammed (1987) is to consider the best linear predictor of ξ, $\hat{\xi}_{BLP}$. It is assumed that the conditional distribution of Y given $X = x$ is normal with mean $\alpha + \beta x$ and variance $\sigma^2_{Y|X}$. Data from a controlled calibration experiment provided estimates of α, β and $\sigma^2_{Y|X}$. It was further assumed that ξ had a distribution with known mean (μ_ξ) and known variance (σ^2_ξ). The author

proposed approximate confidence intervals for ξ based on the estimated M.S.E.($\hat{\xi}_{BLP}$). Using this approach, $\hat{\xi}_c$ and $\hat{\xi}_I$, the classical and inverse estimators respectively, are obtained as special cases of $\hat{\xi}_{BLP}$ and if one assumes that μ_ξ and σ_ξ^2 can be regarded as parameters of a subjective probability distribution for ξ, a full Bayesian analysis is made possible and the resulting theory coincides with that of Hoadley (1970) and Aitchison and Dunsmore (1975). The classical estimator $\hat{\xi}_c$ is obtained when $\sigma_\xi^2 \to \infty$ and the inverse estimator $\hat{\xi}_I$ is obtained when $\mu_\xi = \bar{x}$, $\sigma_\xi^2 = \dfrac{S_{xx}}{(n-2)}$ Other prior assumptions on ξ may justify intervening estimators between $\hat{\xi}_c$ and $\hat{\xi}_I$. The paper should be compared with that of Lwin and Maritz (1982) and also with Brown's (1979) paper. Brown considered the best linear predictor of ξ using integrated mean square error rather than M.S.E. as his optimising criterion.

Naszodi (1978) used traditional propagation-of-error techniques (i.e. estimates of the first two moments obtained from a Taylor expansion) with a bias correction to derive a new estimator of ξ which is practically unbiased, is more efficient than the classical estimator $\hat{\xi}_c$ and is consistent. His expression for the approximate bias of $\hat{\xi}_c$ agrees with that of Shukla (1972). Whereas Shukla assumed that the errors ε_i in model (2.1.1) were normally distributed, Naszodi assumed only that Y was symmetrically distributed about $E(Y)$. Naszodi's estimator is given by

$$\tilde{\xi} = \bar{x} + \frac{(Y' - \bar{Y})}{(\hat{\beta}_1 + \mathrm{Var}(\hat{\beta}_1)/\hat{\beta}_1)}$$

where $\hat{\beta}_1 = S_{xy}/S_{xx}$. Turiel et al. (1982) used a modified form of Naszodi's estimator in both an inverse median estimation problem and an inverse regression problem. The former arises when one wishes to estimate X given a specified median value of Y.

A two-stage sequential approach is considered in two papers by Perng and Tong (1974, 1977). In the first stage (1974 paper) the sequence $\{Y_{1i}\}$ is observed sequentially for the estimation of β_0 and β_1 the intercept and slope parameters respectively. If $\hat{\beta}_1$ is not significantly different from zero, one does not proceed to the second stage and it is concluded that the model is not suitable for estimation of ξ. Otherwise one proceeds to the second stage and observes $\{Y_{2j}\}$ sequentially. When the experiment terminates a fixed width confidence interval can be constructed for ξ.

The final approach to be considered is that of Vecchia et al. (1989). Changes in calibration curves from one time period to the next, caused by drift, often require measuring devices to be recalibrated at frequent intervals. This, however,

require measuring devices to be recalibrated at frequent intervals. This, however, is not always practical. Suppose we take T calibration periods and suppose the T models for the calibration experiments and prediction stages are given respectively by

$$Y_{ij} = f(x_{ij}; \beta_i) + \text{error} \quad i=1,2,\ldots T \quad j=1,2,\ldots n_i$$

$$Y_{ik}' = f(\xi_{ik}; \beta_i) + \text{error} \quad i=1,2,\ldots T \quad k=1,2,\ldots r_i$$

Vecchia et al. considered a random coefficient regression model where the β_i's are regarded as random variables, varying from period to period, but assuming f to be unchanging throughout. The usual practice is to estimate ξ_{ik} *only* using data from the ith calibration period. Vecchia et al. showed that it was more efficient to combine the data from all calibration periods.

3. AN APPROACH USING SIMULATION AND AN IMPLICIT PRIOR

3.1. ASPECTS OF CLASSICAL AND SMOOTHING SPLINE THEORY

The methods and results considered in this chapter and Chapter 4 are based on non-parametric smoothing splines (Schoenberg, 1964; Reinsch, 1967; Wahba, 1975; Craven and Wahba, 1979; Wegman and Wright, 1983; Silverman, 1985 and Eubank, 1988). It would seem expedient therefore to first consider certain aspects of spline theory. Wegman and Wright (1983) state that in the most general setting, a mathematical spline is the solution to a constrained optimisation problem. The nature of the optimisation problem determines the type of spline which results.

3.1.1 Interpolating splines

Suppose observations Y_i $(i=1,2,....n)$ are made at design points $x_i(i=1,2,...n)$. Assume that $x_1 < x_2 < x_3 ... < x_n$ and the x values lie in some finite interval $[a,b]$. Consider the problem of estimating a curve which passes through the points (x_i,Y_i) $i=1,2,...n$, i.e. there is no smoothing of the curve. Such a problem can be expressed as the following constrained optimisation problem: minimise

$$\int_a^b [D^m f(x)]^2 dx \tag{3.1.1}$$

$$\text{subject to} \quad f \in W_m \quad \text{and} \quad f(x_i) = Y_i \qquad i=1,2,...n \tag{3.1.2}$$

Here W_m denotes the space of functions f on $[a,b]$ such that $D^j f$, $j \leq (m-1)$ is absolutely continuous and $D^m f \in \mathcal{L}_2$ where $\mathcal{L}_2$ is the set of measurable square integrable functions on $[a,b]$ and D represents the differentiation operator. The solution, $\hat{f}_I(x)$, is a natural spline of order $2m$ with knots $\{x_i\}$ $i=1,2,...n$ which interpolates the points (x_i,Y_i) $i=1,2,...n$. It is called a natural interpolating spline of order $2m$.

A spline of order p with knots $t_1,t_2,.....t_k$ is defined to be any function of the form

$$f(x) = \sum_{j=0}^{p-1} \theta_j x^j + \sum_{i=1}^{k} \delta_i (x-t_i)_+^{p-1} \tag{3.1.3}$$

for some set of real coefficients $\theta_0,\theta_1,\theta_2....\theta_{p-1},\delta_1,\delta_2,\delta_3....\delta_k$.

The function $(x-t_i)_+^{p-1}$ is defined as follows:

$$(x-t_i)_+^{p-1} = (x-t_i)^{p-1} \qquad x \geq t_i$$
$$= 0 \qquad x < t_i$$

Fig. 3.1 below shows a graph of $(x - t_i)_+^3$ where $t_i = 9.0$.

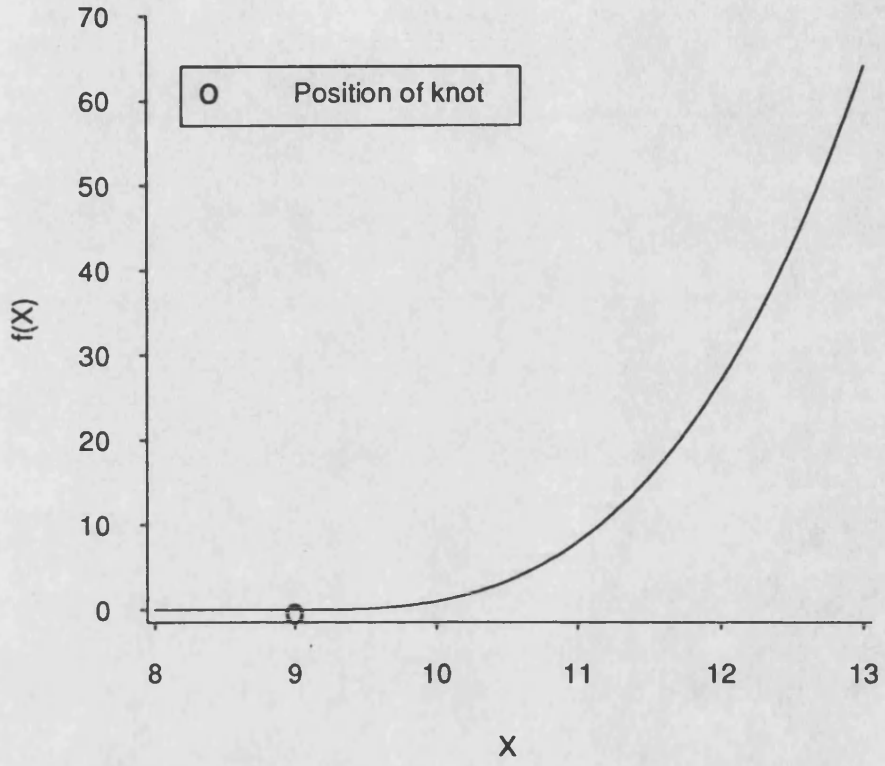


Fig. 3.1: The function $f(x) = (x - t_i)_+^3$

The definition (3.1.3) is equivalent to the following specification :

(a) f is a piecewise polynomial of order p on any sub-interval $[t_i, t_{i+1})$,

(b) f has $(p-2)$ continuous derivatives and

(c) f has a $(p-1)$st derivative that is a step function with jumps at $t_1, t_2, \ldots . t_k$.

(3.1.4)

The *knots* are the points where the different polynomial segments are joined together.

The term *natural* means that in addition to the properties implied by (a), (b) and (c), f is a polynomial of order m outside $[a,b]$. In order for f to be a natural spline we must have $\theta_m = \theta_{m+1} = \ldots = \theta_{p-1} = 0$ since f satisfies the *natural* boundary conditions

$$f^{(j)}(a) = f^{(j)}(b) = 0 \qquad j = m,m+1,\ldots(p-1) \tag{3.1.5}$$

Hence a natural spline of order $2m$ with knots $t_1,t_2,\ldots.t_k$ is a function of the form

$$f(x) = \sum_{j=0}^{m-1} \theta_j x^j + \sum_{i=1}^{k} \delta_i (x-t_i)_+^{2m-1}$$

The case of $m = 2$ (p=4) is of particular interest in this thesis and in this case $\hat{f}_I(x)$ is a natural cubic interpolating spline with knots $\{t_i\}$ (i=1,2,....n). Over $[a,b]$, $\hat{f}_I$ and its first two derivatives are continuous and $\hat{f}_I$ is linear on the two extreme intervals $[a,t_1)$ and $(t_n,b]$. The objective function (3.1.1) represents, in the case of $m = 2$, mean square curvature so $\hat{f}_I$ minimises the mean square curvature of the curve f. The optimal interpolating curve is thus visually smooth.

3.1.2 Smoothing splines

Consider now the more usual statistical problem of where each observation Y_i (i=1,2,...n) is made up of a *signal* plus some random noise. The following model is of interest:

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots.n \tag{3.1.6}$$

where $a < x_1 < x_2,\ldots < x_n < b$, $f \in W_m$ and the random errors ε_i (i=1,2,....n) are assumed to be uncorrelated with zero mean and variance σ^2.

It is desirable to obtain a smooth curve which passes near, in some sense, to the data but is not constrained to interpolate exactly the points (x_i,Y_i) i = 1,2,...n. The estimated curve is the solution to the following constrained optimisation: minimise

$$\sum_{i=1}^{n} (f(x_i)-Y_i)^2 + \alpha \int_a^b [D^m f(x)]^2 dx \tag{3.1.7}$$

$$\text{subject to} \quad f \in W_m \quad \text{and} \quad \alpha > 0 \tag{3.1.8}$$

If one compares this optimisation problem with that given by (3.1.1) and (3.1.2) one will see that the interpolating conditions $f(x_i) = Y_i$ have been replaced by a least squares term in the objective function (3.1.7), which measures the lack of fit of the curve f. As in the objective function (3.1.1), the objective function (3.1.7) contains a penalty term for the lack of smoothness (or degree of roughness) of the curve f, namely

$$\int_a^b [D^m f(x)]^2 dx.$$

The parameter α controls the amount of smoothing and is known as the *smoothing parameter.* If α is too small, the curve will over fit, reducing bias but increasing variance. If α is very large the smoothing term (or roughness penalty term) dominates and not only noise but *signal* may be removed. The choice of α is obviously important. Many authors advocate the method of cross-validation for choosing α and in this and following chapters, the value of α which is chosen in most cases is that value which minimises the asymptotic generalised cross-validation score (Silverman, 1984,1985).

Kimeldorf and Wahba (1970a,b) employed reproducing kernel Hilbert space theory to obtain the solution to the optimisation problem specified by (3.1.7) and (3.1.8). They showed that $\hat{f}_\alpha(x)$, the unique minimiser of (3.1.7), is a natural spline of order $2m$ with knots $\{x_i\}$ $i=1,2,...n$. For the case of $m = 2$, $\hat{f}_\alpha(x)$ is a natural cubic smoothing spline with knots at the data points x_i, $i=1,2,...n$. Eubank (1988) used a mathematically simpler approach to arrive at the same result (see Theorem 5.3, Eubank, 1988). For the case of interest, $m = 2$, $\hat{f}_\alpha(x)$ is a natural cubic smoothing spline called the *spline smoother*, (Greville, 1969, Reinsch, 1967, 1971). As $\alpha \rightarrow 0$, $\hat{f}_\alpha$ converges to the natural cubic interpolating spline, $\hat{f}_I$, discussed in the previous section. As $\alpha \rightarrow \infty$, $\hat{f}_\alpha$ converges to the least squares straight line (the second order polynomial regression estimator of f). This behaviour of $\hat{f}_\alpha$ as α varies from 0 to ∞ is clearly illustrated in Fig. 3.2 which shows a perspective plot of $\hat{f}_\alpha$. The underlying data are the Gompertz calibration data (see Appendix 1 for further details). The number of knots used was eighty and these were equally spaced. A typical coordinate of this three-dimensional plot is $(x_i, \alpha_l, \hat{f}_\alpha(x_i))$ where x_i is the *ith* knot, α_l is the logarithm (base 10) of α, the smoothing parameter and $\hat{f}_\alpha(x_i)$ is the value of the smoothing spline at the knot x_i.

Wegman and Wright (1983) and Silverman (1985) maintain that such non-parametric smoothing splines, $\hat{f}(x)$, provide good estimators of the true function f because

(a) they respond to local variation in the data. For example, the observation at x_i influences only the parts of the spline which are near to it and this influence dies away exponentially.

(b) the amount of smoothing is controllable.

Let $J_m(f) = \int_a^b [D^m f(x)]^2 dx$ and let $\sum_{i=1}^{n} l_i Y_i$ be a linear estimator of f where $l_1, l_2, ...l_n$ are constants which do not involve the Y_i. Li (1982) showed that the minimax estimator of f obtained by minimising

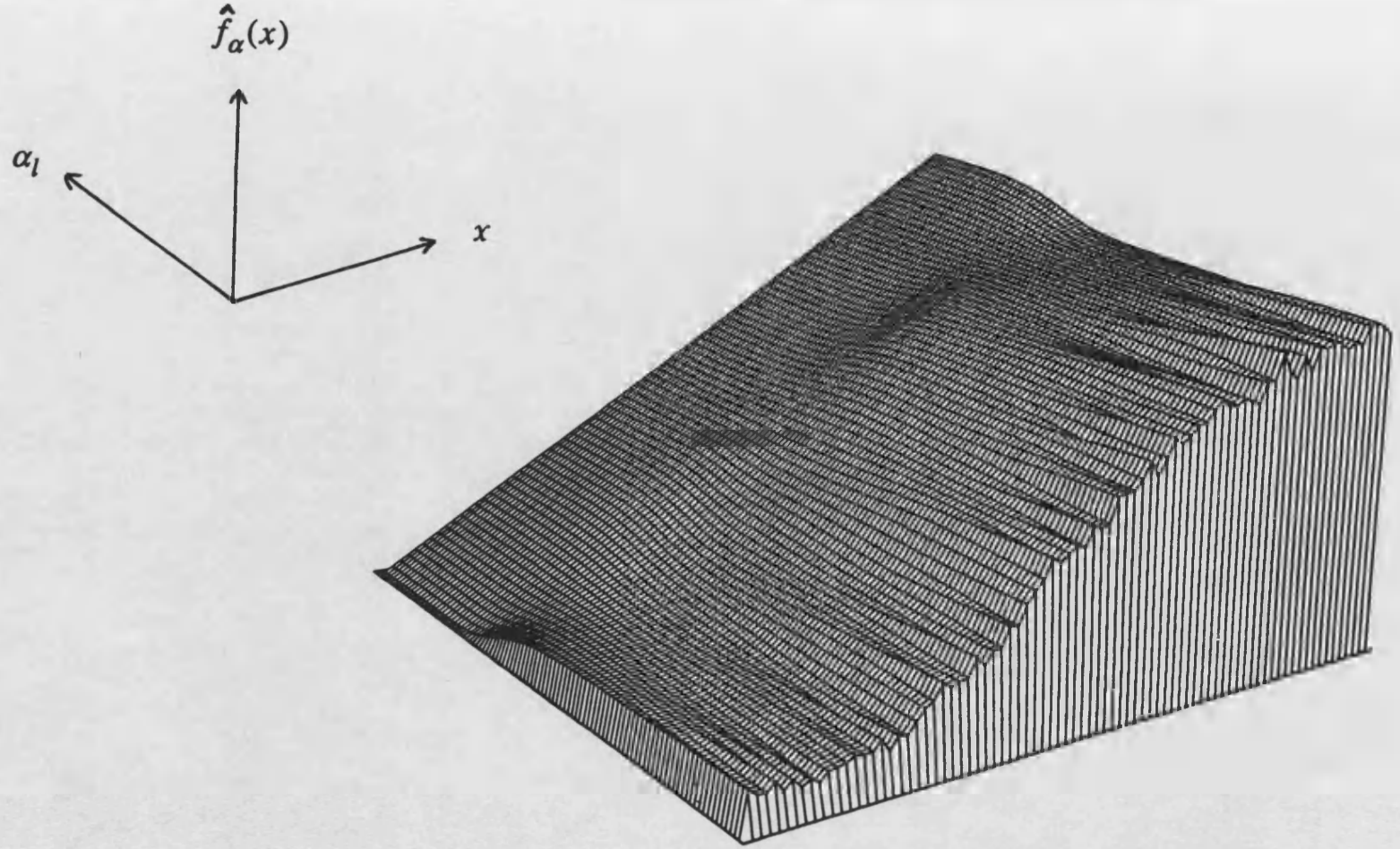


Fig. 3.2: A perspective plot of $\hat{f}_\alpha$

$$\sup_{J_m(f)\leq\rho} E\,[f(x)-\sum_{i=1}^{n} l_i Y_i]^2$$

where $\rho \geq 0$, is the spline smoother $\hat{f}_\alpha(x)$ with $\alpha = \sigma^2/n\rho$. If $f \in W_m$ then $J_m(f) = 0$ if and only if f is a polynomial of order m . This occurs if $\rho = 0$. The spline smoother $\hat{f}_\alpha$ provides a minimax estimator of $f(x)$ for the particular value $\rho = \sigma^2/n\alpha$, so smoothing splines are an extension of polynomial regression which guard against the departure of the calibration curve f from an idealised polynomial regression model. Eubank states that under certain restrictions, a smoothing spline will have a smaller risk than the corresponding mth order polynomial estimator where risk is defined as

$$n^{-1}\sum_{j=1}^{n} E\,[f(x_j)-\hat{f}_\alpha(x_j)]^2$$

Suppose A is an $n \times n$ symmetric matrix which maps the vector of observations Y_i into the vector of predicted values, $\hat{f}(x_i)$, i.e. $\hat{\mathbf{f}} = A\mathbf{Y}$ where

$$\mathbf{Y}^T = (Y_1 Y_2 Y_3 \ldots Y_n) \text{ and } \hat{\mathbf{f}}^T = (\hat{f}(x_1)\hat{f}(x_1)\ldots\hat{f}(x_n))$$

The matrix A is known as the *hat matrix* and is local in character. It depends on α, the smoothing parameter and the design of the points $\{x_i\}$ i=1,2,...n. Also $\hat{f}$ is linear in the observations Y_i since

$$\hat{f}(x_i) = \sum_{j=1}^{n} A_{ij} Y_j \quad i=1,2,\ldots n.$$

If the model (3.1.6) is extended to include the case of random errors ε_i having unequal variances, in particular, $Var\,(\varepsilon_i) = \sigma^2/w_i$ i=1,2,...n where the weights w_i are known, then the constrained optimisation problem becomes: minimise

$$\sum_{i=1}^{n} w_i\,(f(x_i)-Y_i)^2 + \alpha\int_a^b [D^m f(x)]^2 dx \qquad (3.1.9)$$

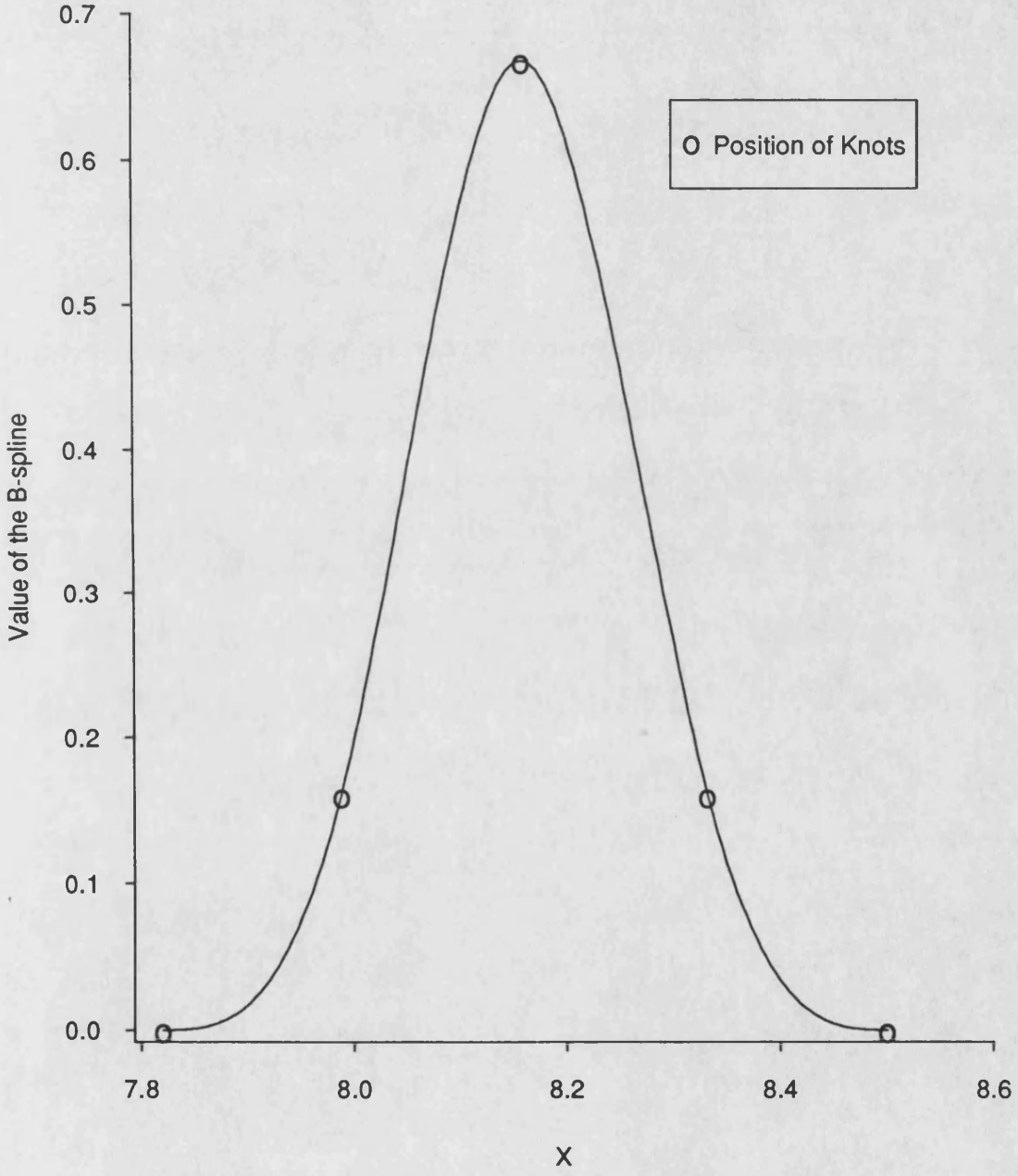
$$\text{subject to} \quad f \in W_m \quad \text{and} \quad \alpha > 0$$

In the case of $m = 2$, the optimal solution is again a natural cubic smoothing spline with knots $\{x_i\}$ i=1,2,...n.

3.1.3 B-splines

The usual definition of a cubic B-spline is a cubic spline with five knots, t_1, t_2, t_3, t_4 and t_5, having the property that it has value zero outside $[t_1, t_5]$. Fig. 3.3 shows a typical cubic B-spline.

Fig.3.3: A typical cubic B-spline



Given data points $\{ x_i \}$ $i=1,2,3.....n$, let Γ be the set of all natural cubic splines with knots $\{ x_i \}$. Then Γ is a vector space of dimension n and we can use B-splines to construct a basis $\{ \beta_1, \beta_2, \beta_3, ..., \beta_n \}$ of Γ which we will call the B-spline basis.

Let us consider first the splines $\beta_3, \beta_4....., \beta_{n-2}$. For $i=3,4,5....,n-2$ β_i is simply a B-spline with knots $x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}$. In the approach of De Boor (1978) and Eubank (1988), there are $(n + 4)$ B-splines denoted by $N_{j,4}(x)$ where $j = -3,-2,-1,...(n-1),n$, constructed by introducing coincident knots $x_{-1} = x_0 = a$ and $x_{n+1} = x_{n+2} = b$. For $i=3,4,5,.....,n-2$, we have

$$\beta_i(x) = N_{i-2,4}(x)$$

A recursive formula (De Boor, 1978) was used to generate these B-splines, $\beta_i(x)$ $i=3,4,5,...,(n-2)$ which is as follows:

$$N_{i,m}(x) = \frac{x-x_i}{x_{i+m-1}-x_i} N_{i,m-1}(x) + \frac{x_{i+m}-x}{x_{i+m}-x_{i+1}} N_{i+1,m-1}(x). \quad (3.1.10)$$

where $$N_{j,1}(x) = \begin{cases} 1 & \text{if } x \in [x_j, x_{j+1}) \\ 0 & \text{otherwise} \end{cases}$$

The end splines $\beta_1, \beta_2, \beta_{n-1}, \beta_n$, (which are not strictly speaking B-splines) were constructed to satisfy the boundary conditions

$$\beta''(x_1) = \beta''(x_n) = 0 \qquad i=1,2,...,n$$

$$\beta'''(a) = \beta'''(b) = 0 \qquad i=1,2,...,n.$$

Note that these conditions already hold for $\beta_3, \beta_4, ..., \beta_{n-2}$. The following equations for $\beta_1, \beta_2, \beta_{n-1}$ and β_n were used:

$$\beta_1(x) = N_{-1,4}(x) + \frac{(x_3+x_2-2a)}{(x_3-a)} N_{-2,4}(x) + \frac{(x_3+x_2+x_1-3a)}{(x_3-a)} N_{-3,4}(x)$$

$$\beta_2(x) = N_{0,4}(x) - \frac{(x_2-a)}{(x_3-a)} N_{-2,4}(x) - \frac{(x_2+x_1-3a)}{(x_3-a)} N_{-3,4}(x)$$

$$\beta_{n-1}(x) = N_{n-3,4}(x) - \frac{(b-x_{n-1})}{(b-x_{n-2})} N_{n-1,4}(x) - \frac{(2b-x_n-x_{n-1})}{(b-x_{n-2})} N_{n,4}(x)$$

$$\beta_n(x) = N_{n-2,4}(x) - \frac{(2b-x_{n-2}-x_{n-1})}{(b-x_{n-2})} N_{n-1,4}(x) + \frac{(3b-x_n-x_{n-1}-x_{n-2})}{(b-x_{n-2})} N_{n,4}(x)$$

(3.1.11)

Our basis will have the property, for all i, that $\beta_i(x) = 0$ if $x \geq x_{i+2}$ or $x \leq x_{i-2}$

for x in the interval $[a, b]$. Given the B-spline basis for Γ, any function f which is a natural cubic spline can be written as a linear combination of the B-splines constituting the basis, i.e.

$$f(x) = \sum_{i=1}^{n} \gamma_i \beta_i(x) \tag{3.1.12}$$

Eubank (1988) obtained the following result (Corollary 5.2) for any basis $\beta_1, \beta_2, \beta_3, \ldots, \beta_n$ of Γ:

Let B and Ω be $n \times n$ matrices defined by

$$B_{ij} = \beta_j(x_i)$$

$$\Omega_{ij} = \int_a^b \beta_i^{(m)}(x) \beta_j^{(m)}(x) dx$$

Provided $n \geq m$, the unique minimiser of (3.1.7) is $\sum_{i=1}^{n} \gamma_i \beta_i$ where $\gamma = (\gamma_1 \gamma_2 \ldots . \gamma_n)^{\mathrm{T}}$ is the solution to

$$(B^{\mathrm{T}}B + \alpha\Omega)\gamma = B^{\mathrm{T}}\mathbf{Y} \tag{3.1.13a}$$

The hat matrix $A(\alpha)$ therefore satisfies the following equation

$$A(\alpha) = B(B^{\mathrm{T}}B + \alpha\Omega)^{-1}B^{\mathrm{T}}. \tag{3.1.13b}$$

Let W be the $n \times n$ diagonal matrix with entries w_i. In the weighted case where the observations Y_i have variances given by $Var(Y_i) = \sigma^2/w_i$ $(i=1,2,\ldots n)$ equations (3.13a) and (3.13b) become

$$(B^{\mathrm{T}}WB + \alpha\Omega)\gamma = B^{\mathrm{T}}W\mathbf{Y} \tag{3.1.13c}$$

and

$$A(\alpha) = B(B^{\mathrm{T}}WB + \alpha\Omega)^{-1}B^{\mathrm{T}}W \tag{3.1.13d}$$

The main advantage of using a B-spline basis is their local support characteristic, namely,

$$\beta_i(x) = 0 \quad \text{if} \quad x \notin [x_{i-2}, x_{i+2}] \quad i=1,2,3,\ldots n$$

This property results in the matrices B and Ω being banded matrices. Matrix B is a band matrix of bandwidth two (i.e. $B_{ij} = 0$ if $|j-i| \geq 2$) and Ω is a band matrix of bandwidth four (i.e. $\Omega_{ij} = 0$ if $|j-i| \geq 4$). The banded nature of these matrices ensures that solutions of equations (3.1.13) are computationally efficient.

3.2 THE POSTERIOR DISTRIBUTIONS OF f AND ξ

The non-parametric Bayesian approach which will be considered in this chapter and Chapter 4 has many similarities with the parametric Bayesian approach of Hunter and Lamboy (1981a) previously discussed in Chapter 2, Section 2.3.

3.2.1 A parametric Bayesian approach

The mathematical models for the calibration process in the parametric case are given in the previous chapter (Equations (2.1.1) and (2.1.2)). The Bayesian approach of Hunter and Lamboy assumes a non-informative prior for (β_0,β_1). The updating of prior beliefs about β_0,β_1 in the light of the calibration experiment data **z** is effected by Bayes theorem to produce a posterior distribution for (β_0,β_1) which is a bivariate Normal distribution. The posterior distribution of (β_0,β_1) effectively defines a posterior distribution for f since

$$f(\xi) = \beta_0 + \beta_1\xi = \eta$$

and f is uniquely determined by β_0 and β_1.

There are two sources of uncertainty in the calibration process. The first is the uncertainty associated with the outcome of the calibration experiment and this is measured by the posterior distribution of (β_0,β_1). The second source of uncertainty is that associated with errors in the future observations $Y_1',Y_2',\ldots,Y_m'$. This source of uncertainty is measured by the posterior distribution of η. The posterior distributions of (β_0,β_1) and η are combined to produce a posterior distribution of ξ using equation (2.1.3). The process is illustrated by a schematic diagram (Fig. 3.4).

3.2.2 A non-parametric Bayesian approach

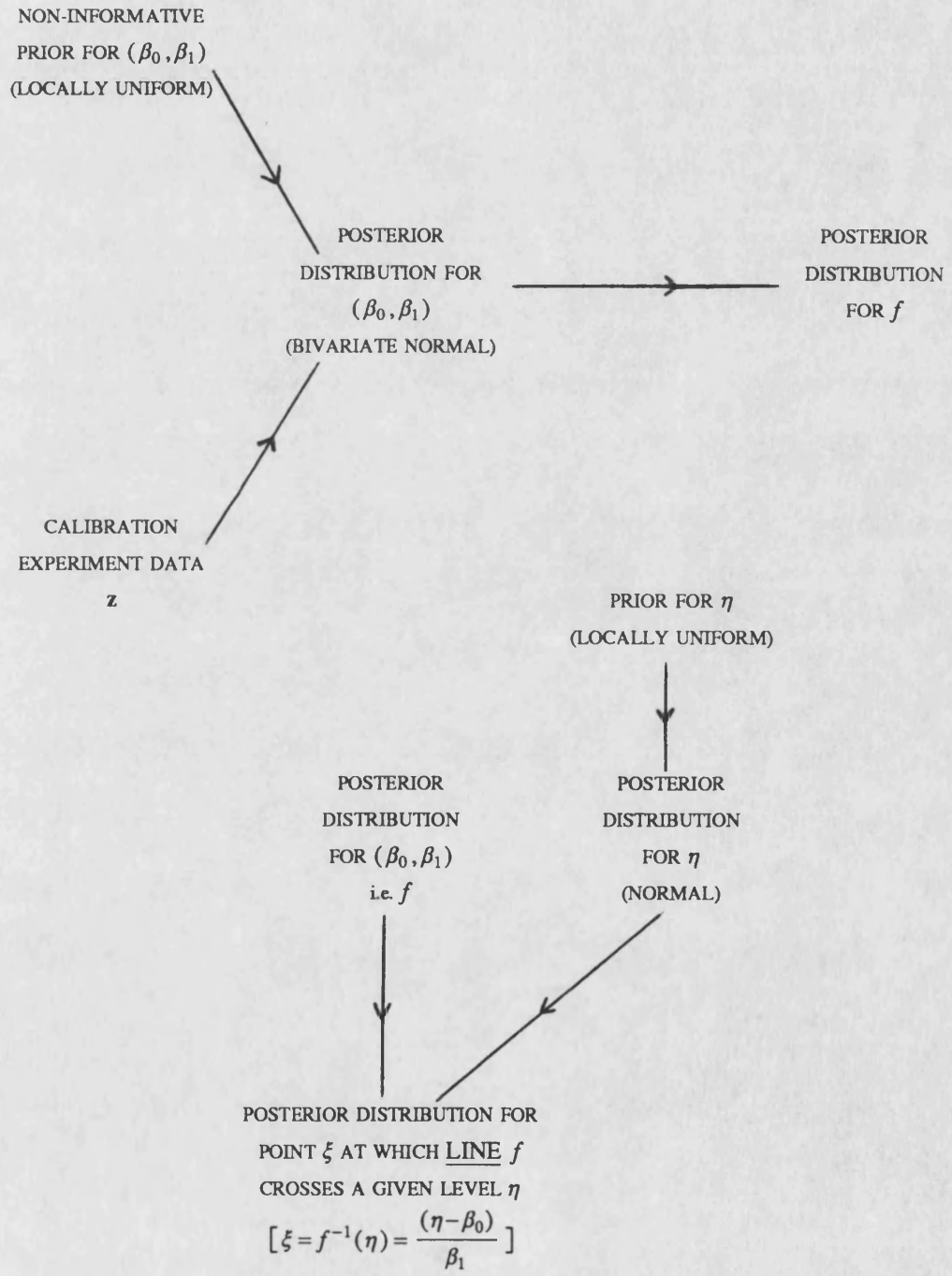
Let us consider the following non-parametric models;

Calibration experiment:

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,n \tag{3.2.1}$$

where $f \in \Gamma$ and the ε_i are assumed independently distributed as $N(0,\sigma^2/w_i)$. This means that f satisfies conditions (3.1.4) and (3.1.5) with $p=4$, $m=2$. Assume for simplicity that all the knots x_i are distinct so that f is a natural cubic spline with n knots $x_1,x_2,x_3,\ldots,x_n$ and $x_1 < x_2 < x_3 \ldots < x_n$.

Fig.3.4 : A parametric Bayesian approach



Prediction stage:

$$Y_j' = f(\xi) + \varepsilon_j' = \eta + \varepsilon_j' \qquad j=1,2,.....m \tag{3.2.2}$$

$$\xi = f^{-1}(\eta). \tag{3.2.3}$$

From equation (3.1.12)

$$f(x) = \sum_{i=1}^{n} \gamma_i \beta_i(x)$$

where β_i is the ith cubic B-spline placed at the knot x_i. The curve f is uniquely determined by $\gamma_1, \gamma_2, ..., \gamma_n$ so it is necessary to provide a prior for γ in the same way that it was necessary to provide a prior for $\beta = (\beta_0\ \beta_1)^T$ in the parametric case. The choice of how much to smooth the data of the calibration experiment corresponds to some sort of prior information. Silverman (1985) proposed a prior distribution for γ of the form:

$$p(\gamma) \propto -\tfrac{1}{2}\lambda \int_a^b [f''(x)]^2 dx \tag{3.2.4}$$

where $\gamma^T = (\gamma_1\ \gamma_2\ \gamma_3\ .....\ \gamma_n)$ and $\lambda = \alpha/\sigma^2$. It will be noted that $\alpha \int_a^b [f''(x)]^2 dx$ is the roughness penalty term in expression (3.1.7) taking $m = 2$. Silverman (1985) showed that given this prior distribution for γ, which corresponds to assuming that γ has a multivariate normal distribution, the posterior distribution of γ is multivariate normal with mean $\hat{\gamma}$ and covariance matrix S^{-1} where

$$S^{-1} = \sigma^2 (B^T W B + \alpha\Omega)^{-1} \tag{3.2.5}$$

$$\hat{\gamma} = \frac{1}{\sigma^2}(S^{-1}B^T W\mathbf{Y}) = (B^T W B + \alpha\Omega)^{-1} B^T W\mathbf{Y} \tag{3.2.6}$$

and B, Ω and W are as defined in Section 3.1.3 above. If one compares the expression for $\hat{\gamma}$ above with equation (3.1.13c) one will see that the spline smoother $\hat{f}_\alpha$ is the posterior mode since the mode equals the mean for an untruncated normal distribution.

It is usual in calibration to assume that the calibration curve is monotonic, in fact it is difficult to know what calibration means in the case of a non-monotonic calibration curve. The true calibration curve is a *monotonic* natural cubic smoothing spline with knots $\{x_i\}$. In order to incorporate this constraint, we assume that f is monotonic increasing or monotonic decreasing. Suppose the model (3.2.1) with this additional assumption is called model (3.2.1a). The prior distribution for γ must be a truncated prior distribution truncated in such a way as to exclude those values of γ which produce non-monotonic splines. Let Θ_{MON} be the set of values of γ which produce monotonic splines. Then the prior distribution of γ is defined as

$$p(\gamma|\gamma \in \Theta_{MON})$$

Let $\mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$ represent a multivariate normal distribution with mean γ and variance matrix S^{-1} which is then truncated to Θ_{MON}. Obviously the mean and variance matrix of the truncated normal distribution will not be $\hat{\gamma}$ and S^{-1} respectively but the mode of the truncated distribution will be $\hat{\gamma}$, provided that $\hat{\gamma} \in \Theta_{MON}$. Using the truncated prior $p(\gamma|\gamma \in \Theta_{MON})$ will produce a truncated posterior distribution for γ. There are two possible approaches here;

(a) Truncate the prior for γ given by expression (3.2.4) to exclude those values of γ which are not in Θ_{MON} and obtain the truncated posterior distribution of γ.

(b) Use the untruncated prior for γ given by expression (3.2.4), obtain the untruncated posterior distribution for γ which is a multivariate normal distribution with mean $\hat{\gamma}$ and variance matrix S^{-1} and then truncate to Θ_{MON}.

If one adopts approach (a), the statistical theory is rather intractable so approach (b) was adopted. The posterior distribution of γ is therefore

$$\mathcal{N}_{MON}(\hat{\gamma}, S^{-1}) \tag{3.2.7}$$

where $\hat{\gamma}$ and S^{-1} are given by equations (3.2.6) and (3.2.5) respectively.

It will be seen from equation (3.2.3) that ξ is a non-linear functional of the curve f (A functional $\psi(f)$ is a mapping from the space of curves to the real numbers). A functional ψ is called linear if, given any curves $f_1, f_2 \in \Gamma$ and real numbers a_1, a_2,

$$\psi(a_1 f_1 + a_2 f_2) = a_1 \psi(f_1) + a_2 \psi(f_2).$$

Let $f_3 \in \Gamma$ and suppose $f_3 = a_1 f_1 + a_2 f_2$ then it can be clearly seen that

$$f_3^{-1}(\eta) \neq a_1 f_1^{-1}(\eta) + a_2 f_2^{-1}(\eta)$$

for any fixed η so since $\xi = f^{-1}(\eta)$, ξ is a non-linear functional of f. The posterior distribution of ξ is not tractable since it is a non-linear function of a high-dimensional truncated multivariate normal distribution, so the question arises as to how to obtain the posterior distribution of ξ. This is achieved by simulation.

The inverse covariance matrix S^{-1}, is a symmetric band matrix of bandwidth four and possesses a Cholesky decomposition given by

$$S = LL^{\mathrm{T}}$$

where L is a lower-triangular band matrix of bandwidth four. The first step of the procedure for obtaining the posterior distribution of ξ is to simulate from the

procedure for obtaining the posterior distribution of ξ is to simulate from the posterior distribution of f which Silverman (1985) demonstrated to be easy. Let

$$L^{\mathrm{T}}\gamma = L^{\mathrm{T}}\hat{\gamma} + \mathbf{z} \tag{3.2.8}$$

where L^{T} is given by $LL^{\mathrm{T}} = (B^{\mathrm{T}}WB+\alpha\Omega)/\sigma^2$, $\hat{\gamma}$ is as defined in equation (3.2.6) and $\mathbf{z}$ is an $n\times 1$ vector of independent $N(0,1)$ random variables. Each posterior realisation of γ and hence each posterior realisation of f involved solving the set of equations (3.2.8) which are a band-limited upper triangular linear system of equations. Fig. 3.5 shows three posterior realisations of f for the Gompertz data set. Only part of the vertical scale is shown on the graph so that the three realisations can be clearly distinguished from each other.

As mentioned above, the posterior distribution of γ is truncated to Θ_{MON}, so the posterior distribution of f will be truncated to exclude non-monotonic splines. Hence any realisations of f which were non-monotonic were rejected. It was easy to achieve this because of the polynomial nature of f within each subinterval $[x_i,x_{i+1})$ $i=1,2,\ldots,n-1$. Suppose it is assumed that f is a monotonic increasing function of x, and further suppose that within the interval $[x_i,x_{i+1})$ f is defined as follows:

$$f(x) = a_i(x-x_i)^3 + b_i(x-x_i)^2 + c_i(x-x_i) + d_i.$$

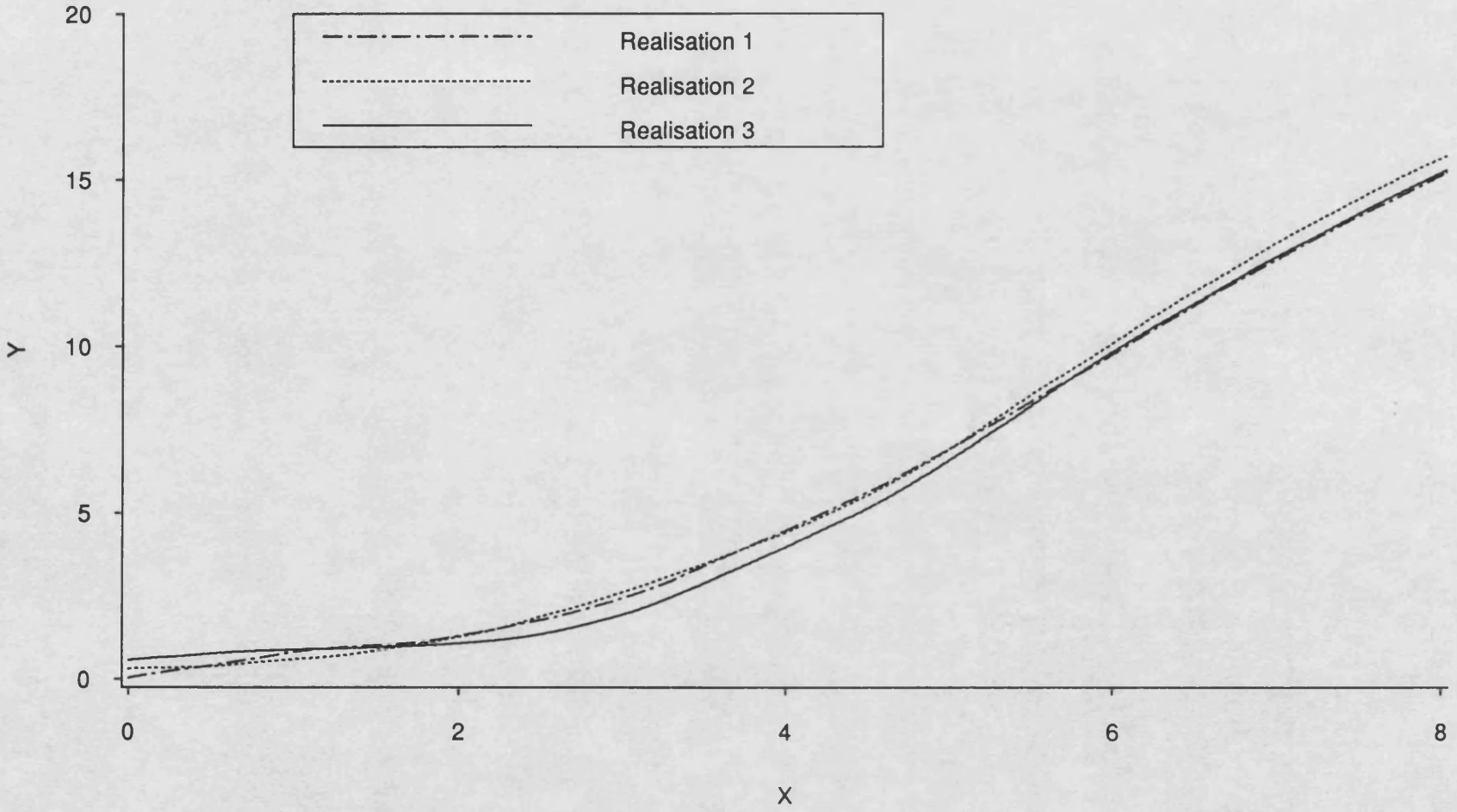
The gradient of the curve, $f'(x)$, is a quadratic function of x given by

$$f'(x) = 3a_i(x-x_i)^2 + 2b_i(x-x_i) + c_i.$$

It is necessary that $f'(x) \geq 0$ for monotonicity and it can be shown that $f'(x)$ becomes negative if either $c_i < 0$ or in the case of $a_i > 0$ and $c_i > 0$, the larger real root of the equation $f'(x) = 0$ is less than h_i where $h_i = (x_{i+1}-x_i)$. For each of the data sets considered in subsequent sections, 200 posterior realisations of f were generated. Suppose these are denoted by $\tilde{f}_1,\tilde{f}_2,\tilde{f}_3,\ldots\ldots\tilde{f}_{200}$.

The next stage in the procedure for obtaining the posterior distribution of the non-linear functional ξ was to combine the posterior distributions of f and η. Hunter and Lamboy (1981a) assumed a locally uniform prior for η. When these prior beliefs about η were updated in the light of the m future observations $Y_1',Y_2',Y_3',\ldots,Y_m'$, it produced a posterior distribution for η which was normal with mean $\bar{Y}'$ and variance σ^2/m where $\bar{Y}' = \sum_{j=1}^{m} Y_j'$. It was decided to assume, in the non-parametric case, a uniform prior for η on $[\tilde{f}(x_1), \tilde{f}(x_n)]$ where $\tilde{f}$ is a posterior realisation of f. The resulting posterior distribution of η is a truncated normal distribution defined as follows:

Fig.3.5: Posterior realisations of f (Gompertz data set)



$$p(\eta|\mathbf{Y}') = (\frac{2\pi\sigma^2}{m})^{-\frac{1}{2}} \exp[\frac{-m(\eta-\bar{Y}')^2}{2\sigma^2}] \qquad \tilde{f}(x_1) \leq \eta \leq \tilde{f}(x_n)$$

$$= 0 \qquad \text{otherwise.}$$

For each data set considered, 50 posterior realisations of η were generated for each posterior realisation of f and the values of

$$\tilde{f}_s^{-1}(\eta_v) \qquad \begin{matrix} s = 1,2,\ldots,200 \\ v = 1,2,\ldots,50 \end{matrix}$$

were calculated producing 10,000 posterior realisations of ξ. (A *different* set of posterior realisations η_v were generated for *each* posterior realisation of f, $\tilde{f}_s$).

Although some of the posterior distributions were approximately symmetrical others were skewed. It was therefore decided to use the posterior median as a point estimator of ξ . To facilitate the calculation of this, the last posterior realisation of ξ was ignored giving a posterior distribution consisting of 9999 realisations of ξ .

The 100pth quantile of a continuous probability distribution with distribution function $F(x)$ is that number x_p such that

$$F(x_p) = p$$

Let $X_1, X_2, X_3 \ldots . X_n$ be a random sample of a continuous random variable with distribution function $F(x)$ and let $X_{(1)}, X_{(2)}, X_{(3)} \ldots . X_{(n)}$, represent the order statistics. Then $U_{(j)} = F(X_{(j)})$ $j=1,2,\ldots,n$ are the order statistics for a uniform (0,1) random variable and $U_{(j)}$ has a Beta distribution with parameters j and $n-j+1$. Therefore

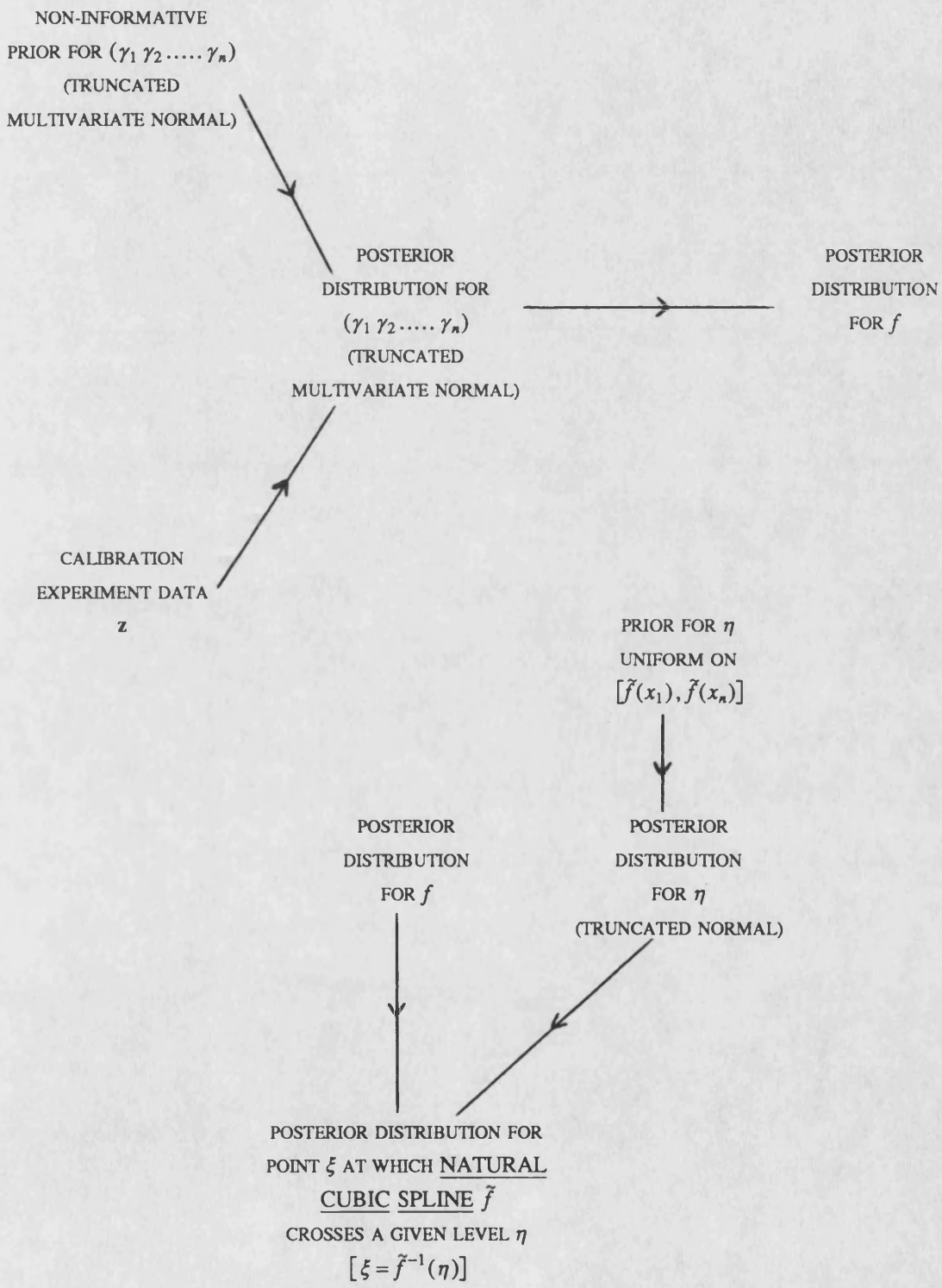
$$E[F(X_{(j)})] = j/(n+1)$$

Then with $p = j/(n+1)$ a natural estimator of x_p is $X_{(j)}$. Here $n+1 = 10,000$ so the point estimator of the posterior median is $X_{(5000)}$. Symmetrical probability intervals for ξ were constructed using the relevant order statistics . In particular a 90% probability interval for ξ was constructed by using

$$[X_{(500)}, X_{(9500)}]$$

A schematic diagram showing all the steps of this non-parametric approach is given in Figure 3.6. Comparison of Figures 3.4 and 3.6 reveal many similarities.

Fig.3.6 : A non-parametric Bayesian approach



3.3 APPLICATION OF METHOD TO VARIOUS GROUPS OF DATA SETS

3.3.1 Simulated data sets

Six simulated data sets were analysed, (five of them, (a) to (e), arising from non-linear models) namely,

(a) Gompertz data set

(b) Weibull data set

(c) Preece-Baines data set

(d) Bleasdale-Nelder data set

(e) Asymptotic data set

(f) Linear data set

Full details of the data sets are given in Appendix 1. For all six calibration data sets analysed, there were eighty evenly spaced knots covering the calibration range and three observations simulated at each knot value giving a total of 240 observations. The model for the calibration experiment is therefore given by

$$Y_{ik} = f(x_i) + \varepsilon_{ik} \qquad \begin{matrix} i=1,2,\ldots 80 \\ k=1,2,3 \end{matrix}$$

where f is a monotonic natural cubic smoothing spline with knots $\{x_i\}$ i=1,2,...80 and the ε_{ik} are assumed independently distributed as $N(0,\sigma^2)$. The posterior distribution of γ is a truncated multivariate normal given by (3.2.7) i.e.

$$\mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$$

where $\hat{\gamma}$ and S^{-1} are given by

$$\begin{aligned} S^{-1} &= \sigma^2 (B^{\mathrm{T}}WB+\alpha\Omega)^{-1} \\ \hat{\gamma} &= (B^{\mathrm{T}}WB+\alpha\Omega)^{-1}B^{\mathrm{T}}W\mathbf{Y} \end{aligned} \qquad (3.3.1)$$

Here W is an 80×80 diagonal matrix with entries $w_i = 3$ and $\mathbf{Y}$ is a vector (80×1) of mean values, i.e. $\mathbf{Y}^{\mathrm{T}} = (\bar{Y}_1\bar{Y}_2\ldots\bar{Y}_{80})$ where $\bar{Y}_i$ is the mean of the observations at knot i (i=1,2,...80).

Figures 3.7 - 3.12 show the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder, Asymptotic and Linear calibration data sets respectively and the corresponding spline smoothers $\hat{f}_\alpha(x)$. Fig. 3.12 also shows the least squares line. The values of the smoothing parameter, α, were obtained by using the computer package BATHSPLINE (Silverman and Watters, 1984) which seeks to minimise the asymptotic generalised cross-validation score (Silverman, 1984).

Fig.3.7: Gompertz data set

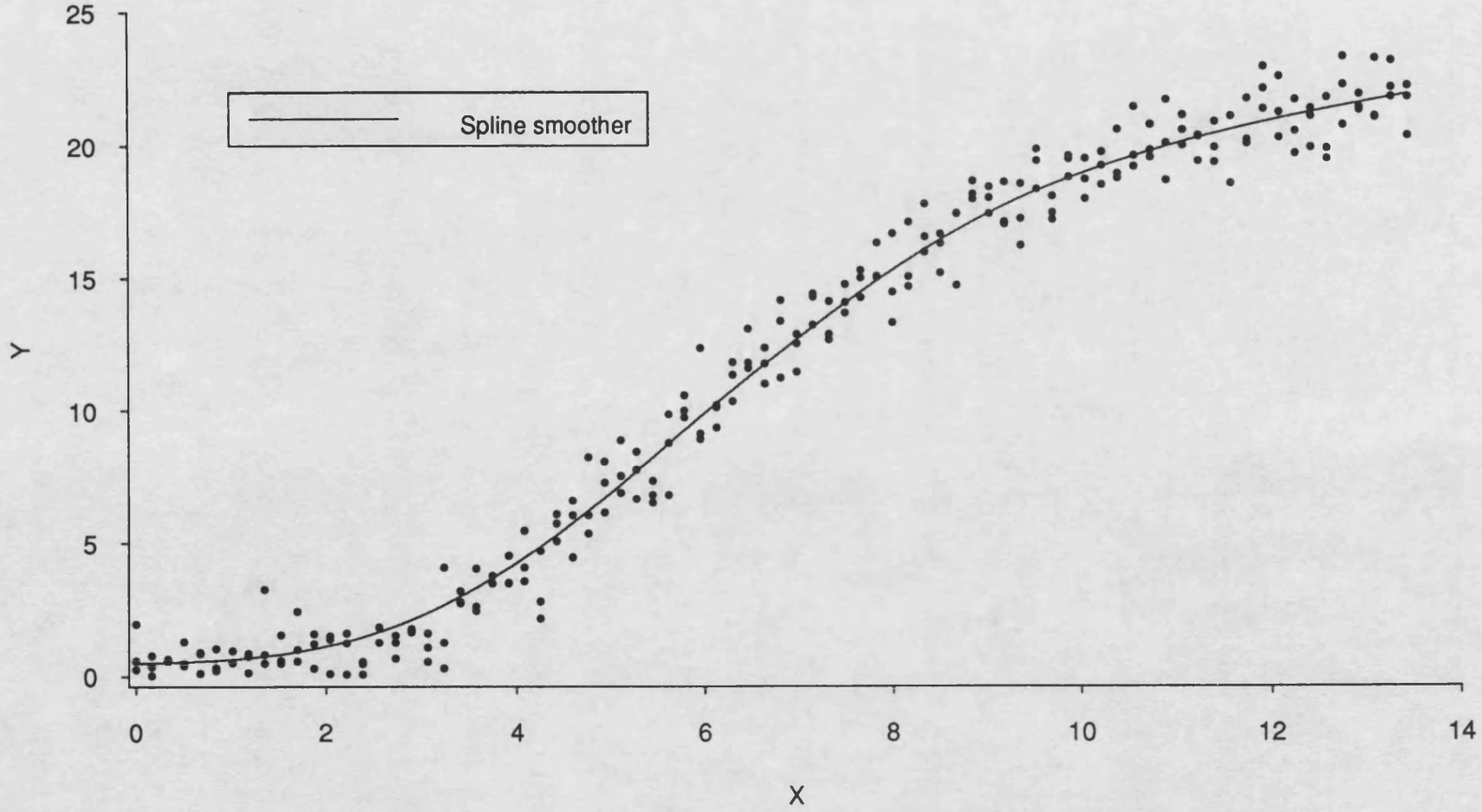


Fig.3.8: Weibull data set

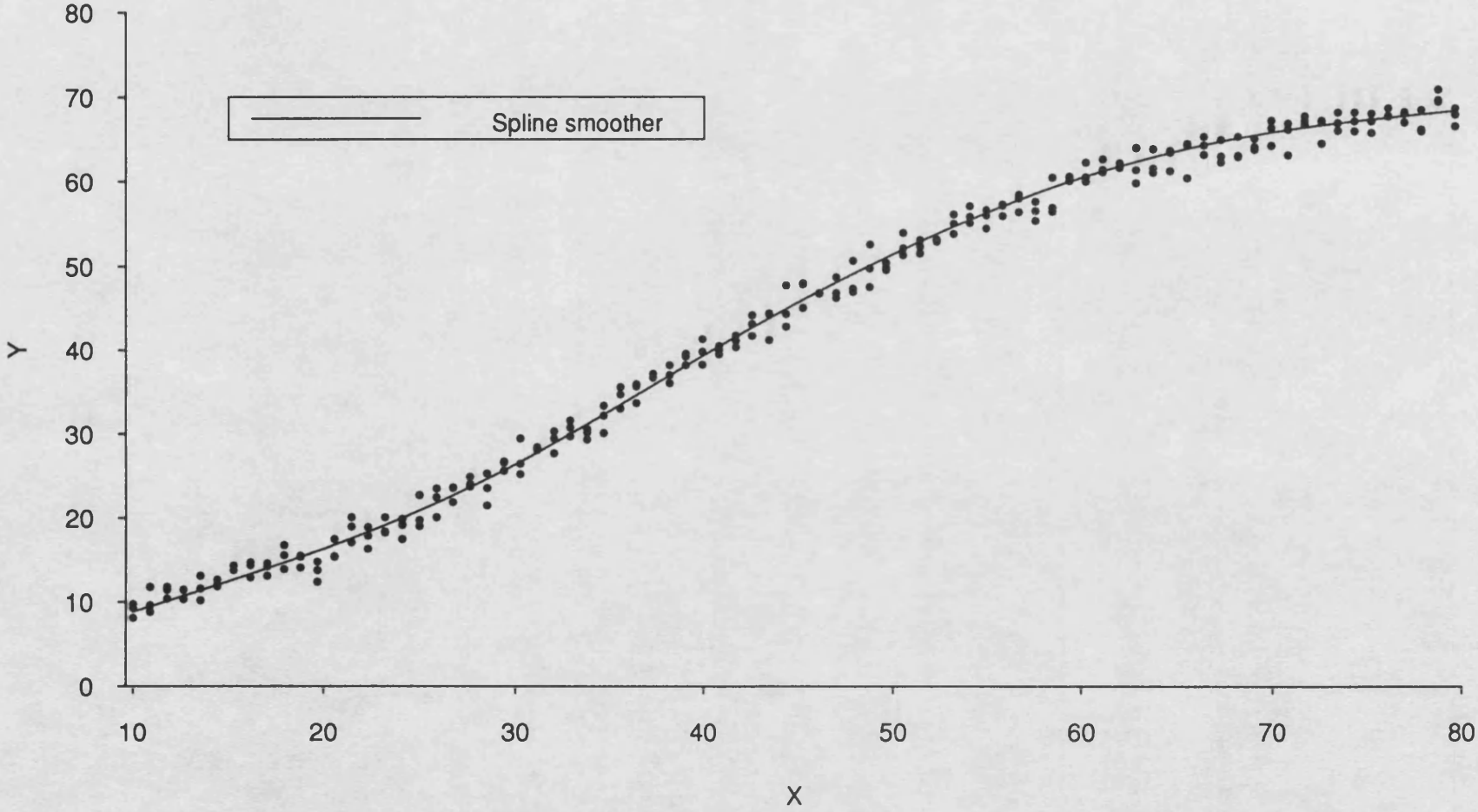


Fig.3.9: Preece-Baines data set

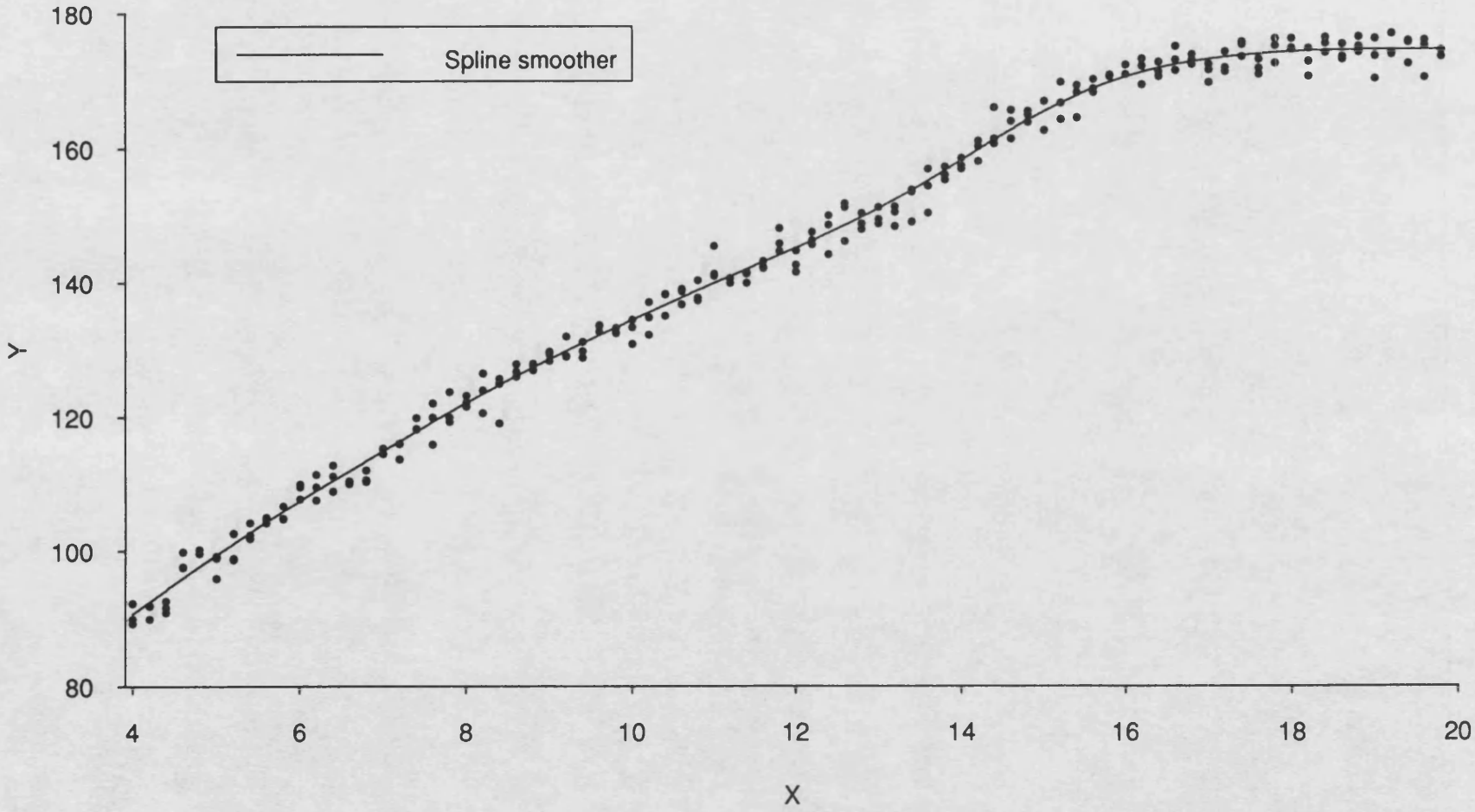


Fig.3.10: Bleasdale-Nelder data set

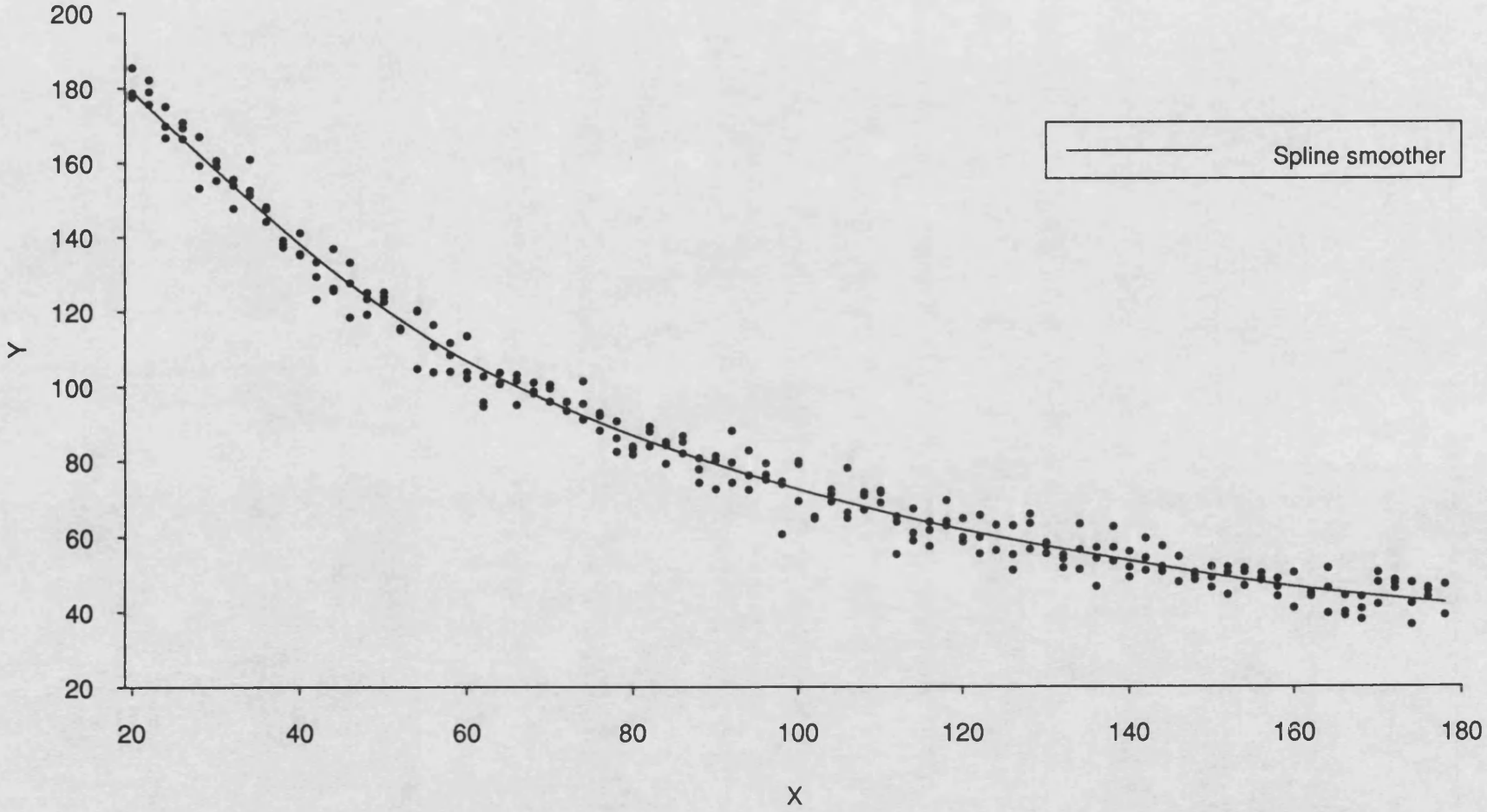


Fig.3.11: Asymptotic data set

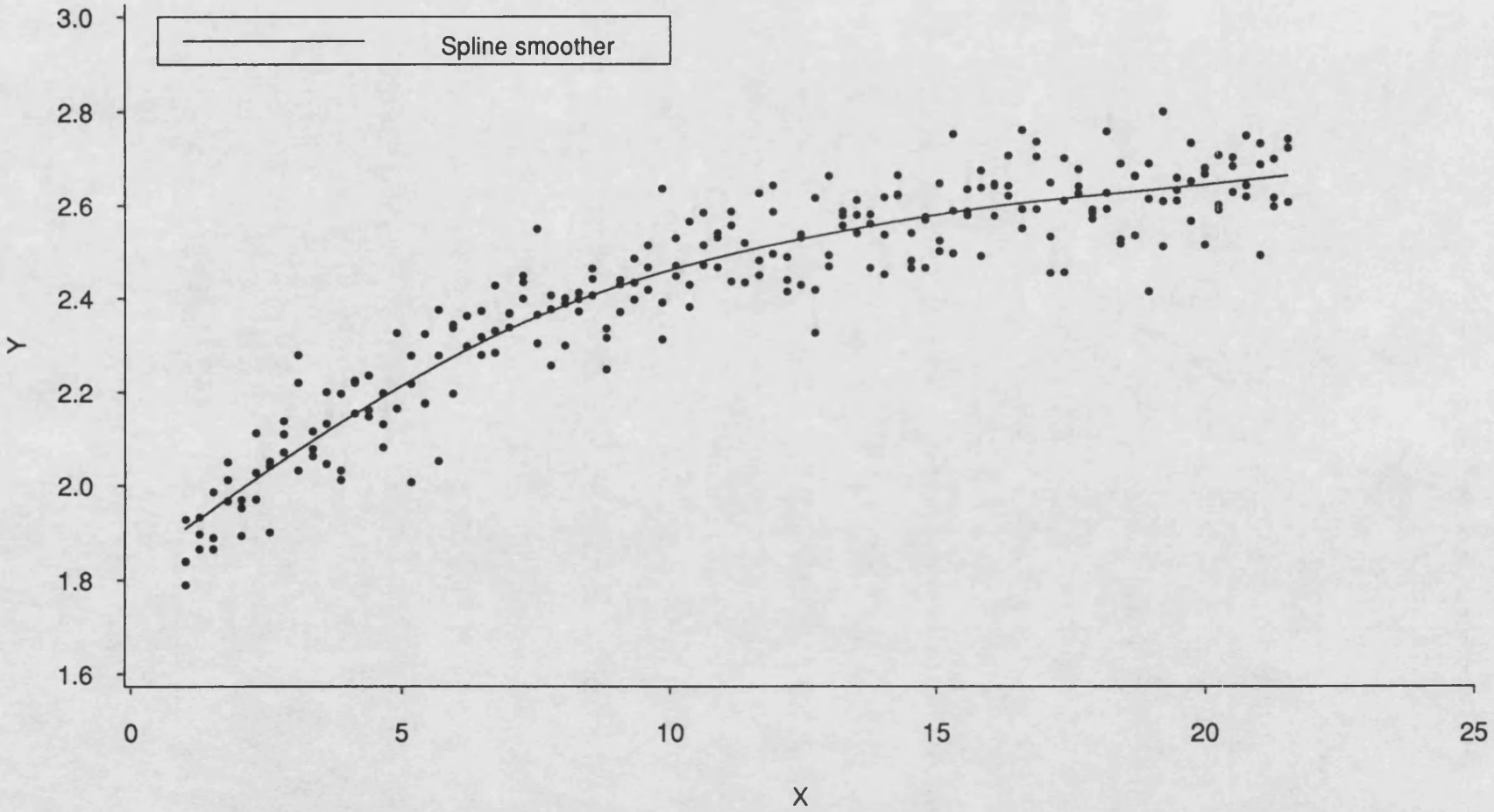
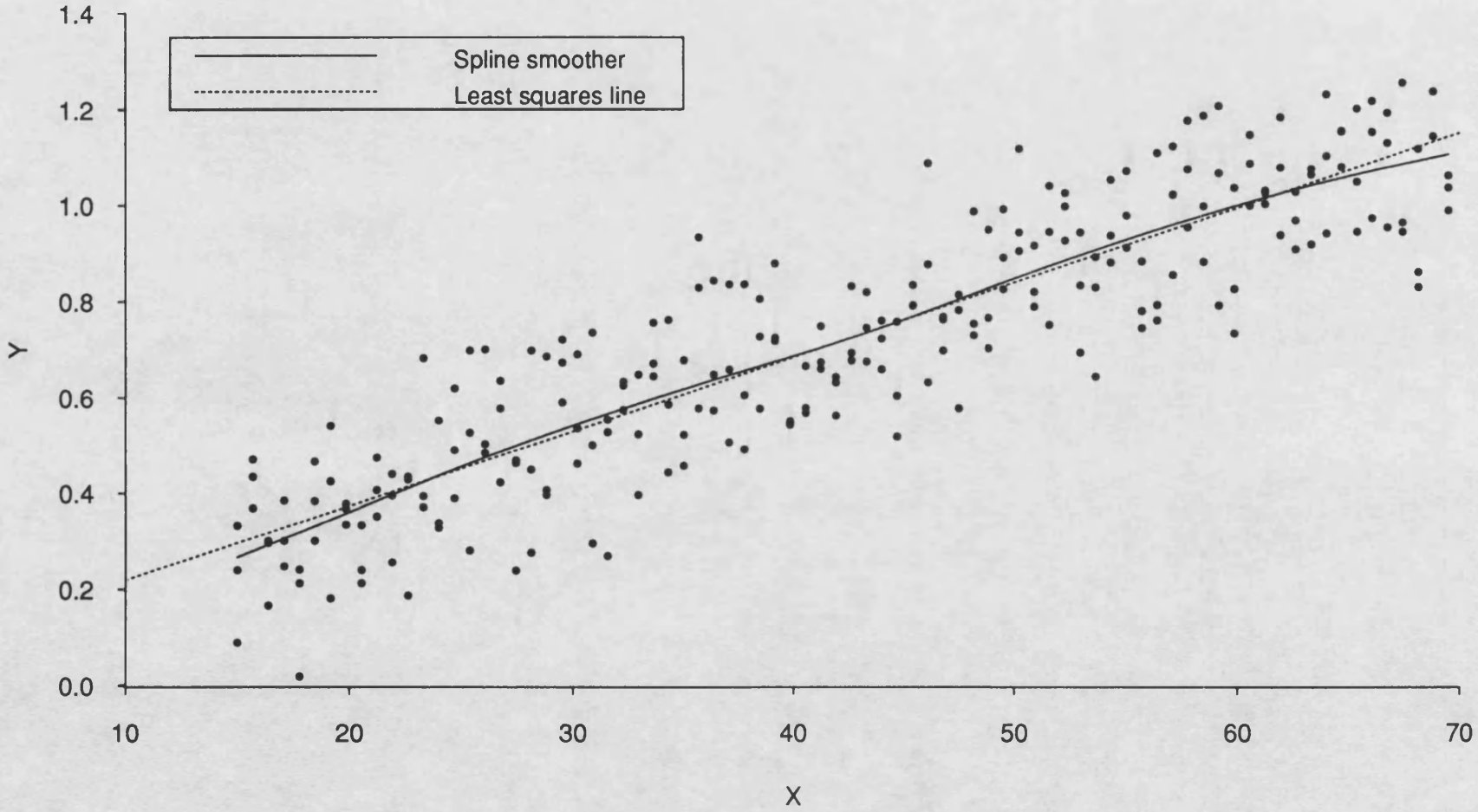


Fig.3.12: Linear data set



To assess the performance of the method at the prediction stage of the calibration process, three observations were simulated at each of twenty newly chosen values of X covering the calibration range so that the model for the prediction stage is given by equation (3.2.2) with $m = 3$, i.e.

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3.$$

In the model, ξ is usually the *unknown* value of X corresponding to the observations Y_1', Y_2', Y_3'. However, to enable assessment of the method, the twenty values of ξ were known. Posterior distributions of ξ were obtained using the method described in Section 3.2.2 and the posterior median used as a point estimate of ξ. Suppose the point estimate is denoted as $\hat{\xi}$. Table 3.1 gives the mean absolute error for each of the six data sets together with the value of the smoothing parameter, α, and the calibration range. Here absolute error is defined as $|\xi - \hat{\xi}|$.

In Fig. 3.13, the estimated value $\hat{\xi}$ is plotted against the true value ξ for the Gompertz data set. The 45° line represents zero error and it can be seen that errors are small. Also the larger errors tend to coincide with the parts of the spline which are rather flat, i.e. the gradient is small or close to zero. With the Gompertz calibration data set, the slope of the spline smoother (see Fig. 3.7) is small over the range $0 \leq X \leq 3$ and $X > 11$ and this is reflected in the errors. This is a good feature of the method because where the spline is rather flat, the calibration data are not very informative concerning the value of ξ and this should be reflected in the estimates of ξ. Fig. 3.14 shows the estimated value $\hat{\xi}$ plotted against the true value ξ for the Weibull data set in the case of $m = 1$ and $m = 3$. It will be noted that even when $m = 1$, i.e. only one observation Y' at the unknown ξ, the errors are still quite small.

Interval estimates were obtained for ξ and the estimate was termed *successful* if the true value ξ was contained within the calculated interval. In particular, 90% interval estimates were obtained using the order statistics as discussed in Section 3.2.2, namely,

$$[X_{(500)}\ ,\ X_{(9500)}]$$

Table 3.2 gives details of the interval estimates for each of the six data sets analysed. In Table 3.2, interval estimates for $m = 1$, $m = 2$ and $m = 3$ are given for the Weibull data set. As expected, the larger m is, the more precise are the interval estimates.

In Chapter 7, both point and interval estimates for ξ are given for the Linear data set using the parametric approach of Hunter and Lamboy (1981a).

Table 3.1 Point estimates for ξ (simulated data)

Data set	Calibration range	Value of smoothing parameter	Mean absolute error
Gompertz	0.00 - 13.43	10	0.39
Weibull			
m=1 in eqn. (3.2.2)			1.64
m=2 in eqn. (3.2.2)	10.00 - 79.52	1395	0.88
m=3 in eqn. (3.2.2)			0.72
Preece-Baines	4.00 - 19.80	4	0.11
Bleasdale-Nelder	20.00 - 178.00	15567	3.15
Asymptotic	1.00 - 21.74	140	1.04
Linear	15.00 - 70.00	2613	3.57

Fig.3.13: Comparison of point estimates
(Gompertz data set)

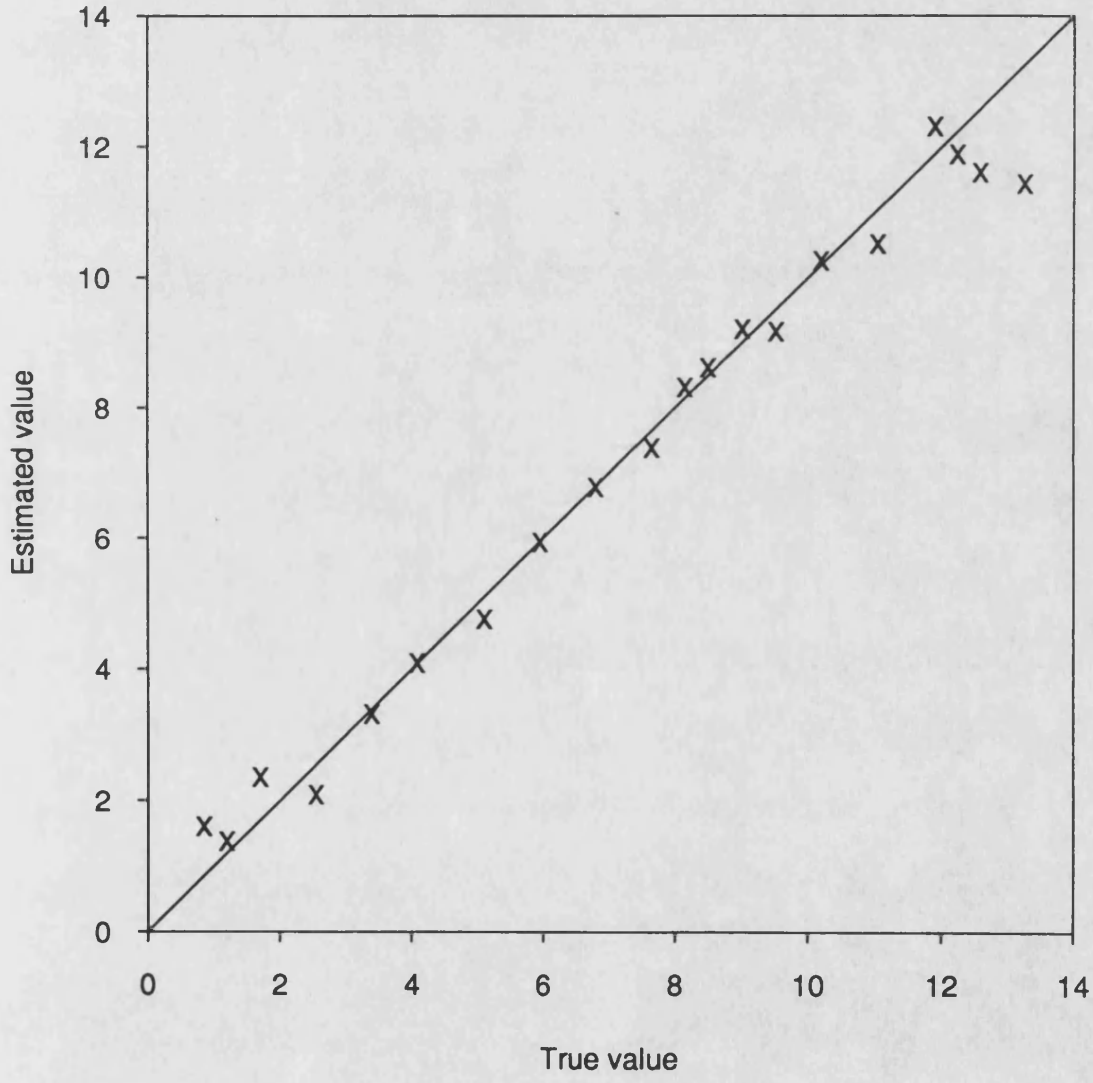


Fig.3.14: Comparison of point estimates
(Weibull data set)

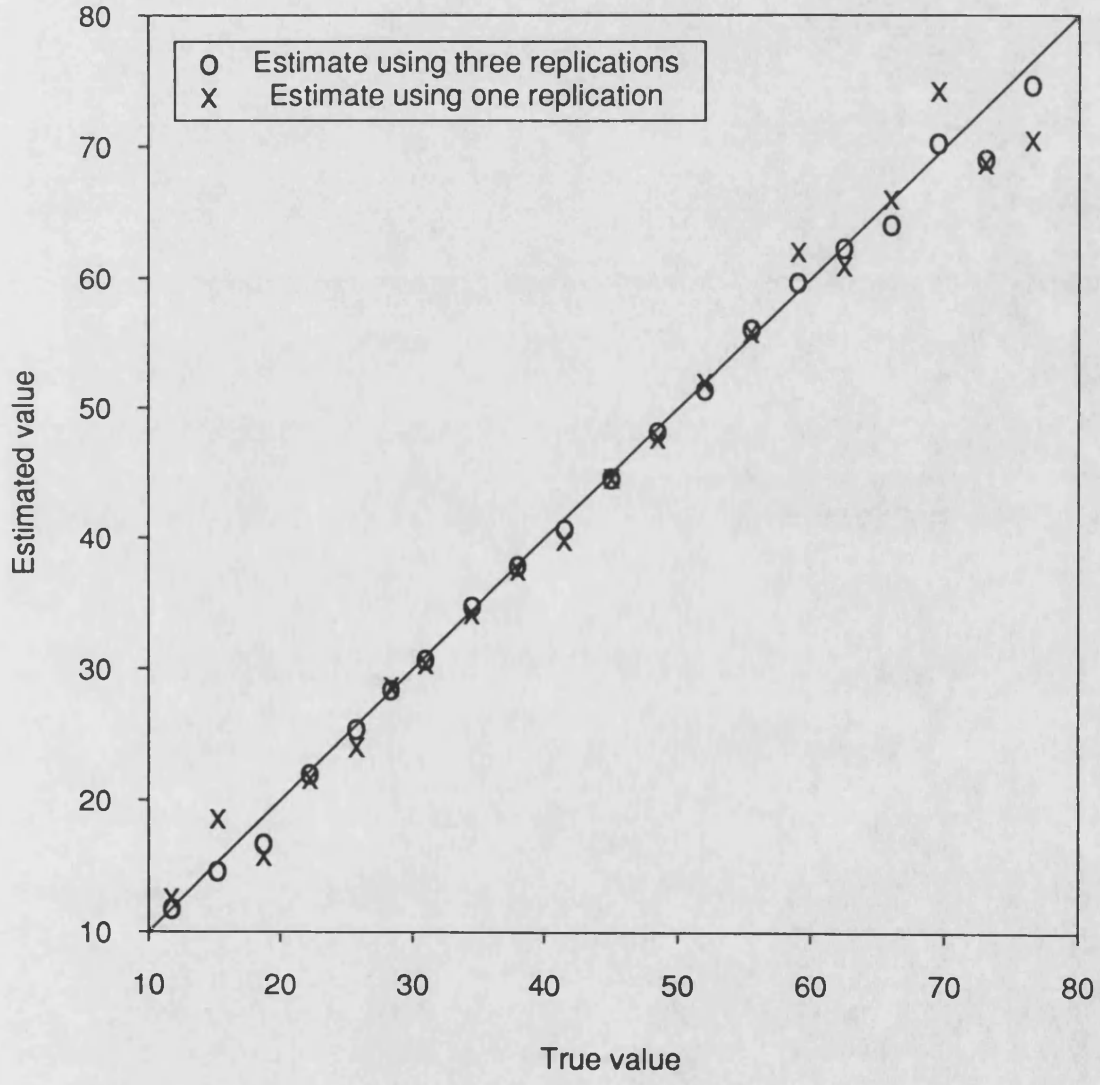


Table 3.2 Interval estimates for ξ (simulated data)

Data set	Max. interval width	Min. interval width	Mean interval width	Proportion of successful estimates
Gompertz	2.43	0.68	1.56	19/20
Weibull				
m=1 in model (3.2.2)	11.71	3.22	5.80	17/20
m=2 in model (3.2.2)	9.82	2.30	4.22	20/20
m=3 in model (3.2.2)	8.06	1.91	3.53	19/20
Preece-Baines	2.92	0.17	0.72	19/20
Bleasdale-Nelder	26.25	4.05	13.14	18/20
Asymptotic	8.50	1.18	4.96	19/20
Linear	16.78	10.11	15.12	18/20

These are compared there with the corresponding results when using the non-parametric approach detailed above. Also in Chapter 7, Hunter and Lamboy's approach is applied to the Weibull data and the resulting point and interval estimates for ξ compared with those obtained using the non-parametric approach described in this chapter.

3.3.2 Data on length of transparent root dentine

The data considered in this section consisted of measurements of intact tooth transparency (ITTM) made on the intact teeth extracted from 43 patients with known ages. The number of teeth obtained from each patient ranged from 1 to 12. The 153 teeth examined, were classified into six types; upper central, upper lateral, upper canine, lower central, lower lateral and lower canine. The upper teeth came from the upper jaw and the lower teeth from the lower jaw. The position in the mouth, right or left, was recorded. Obviously measurements on the right and left teeth from the *same* patient are correlated and this was taken into account in the analysis. Further details of the data sets are given in Appendix 1. Graphs of ITTM against age for all six types of teeth showed a tendency for the variability to increase with age so the data were transformed with a logarithmic transformation to remove the non-homogeneity of the variability.

Suppose there are m patients with ages $x_1, x_2, \ldots, x_m$ each contributing one or two teeth of a particular type. For any particular type of tooth, the mathematical model for the calibration experiment is the following mixed model:

$$Y_{ik} = f(x_i) + \tau_i + \eta_{ik} \qquad \begin{matrix} i=1,2,\ldots.m \\ k=1 \text{ or } 2 \end{matrix} \qquad (3.3.2)$$

where Y_{ik} = log ITTM for the kth tooth of the ith patient, τ_i is the effect due to the ith patient and f is a monotonic natural cubic smoothing spline with knots $\{x_i\}$ $i=1,2,\ldots m$. It is assumed that the τ_i are i.i.d. normal random variables with mean zero and variance σ_1^2, and the η_{ik} are i.i.d. normal random variables with mean zero and variance σ_2^2. It is further assumed that the η_{ik} are independent of the τ_i.

Using the model defined in equation (3.3.2)

$$\text{Cov}\,(Y_{ik}, Y_{ik'}) = \begin{cases} \sigma_1^2 & k \neq k' \\ \sigma_1^2 + \sigma_2^2 & k = k' \end{cases}$$

and Cov $(Y_{ik}, Y_{i'k'}) = 0$ if $i \neq i'$.

Suppose the correlation between the right and left teeth of a particular type taken from a particular patient is measured by the correlation coefficient ρ then

$$\rho = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma_1^2}{\sigma^2}$$

where $\mathrm{Var}(Y_{ik}) = \sigma_1^2+\sigma_2^2 = \sigma^2$.

The correlation coefficient ρ is the same for different patients providing two teeth of a particular type. Let $\varepsilon_{ik} = \tau_i + \eta_{ik}$ then the model (3.3.2) can be written as follows:

$$Y_{ik} = f(x_i) + \varepsilon_{ik} \qquad \begin{matrix} i=1,2,\ldots,m \\ k=1 \text{ or } 2 \end{matrix}$$

where the ε_{ik} are normally distributed random variables with mean $\mathbf{0}$ and covariance matrix V where

$$V = \begin{cases} \begin{bmatrix} \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 \end{bmatrix} & k = 2 \\ \begin{bmatrix} \sigma_1^2 + \sigma_2^2 \end{bmatrix} & k = 1 \end{cases}$$

It may be recalled from Section 3.2.2 that the posterior distribution of γ is a multivariate normal truncated to exclude those values of γ which produce non-monotonic splines. Using the notation of Section 3.2.2. the posterior distribution of γ is truncated to Θ_{MON}, in particular,

$$\mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$$

where S and $\hat{\gamma}$ are as given by equations (3.3.1). Suppose for simplicity that all the m ages $x_1, x_2 \ldots . x_m$ are distinct, then for the case of $k = 2$,

$$\mathrm{Var}(\bar{Y}_{i\cdot}) = \mathrm{Var}(\frac{Y_{i1}+Y_{i2}}{2}) = \tfrac{1}{2}\sigma^2(1+\rho)$$

where $\bar{Y}_{i\cdot}$ is the mean of the observations at knot i (i=1,2,...m). So the weight attached to knot i when $k = 2$ is $2/(1+\rho)$. For the case of $k = 1$, $\mathrm{Var}(\bar{Y}_{i\cdot}) = \sigma^2$ so the weight attached to knot i when $k = 1$ is 1. Hence the diagonal matrix W in equations (3.3.1) is an $m \times m$ matrix with entries w_i where

$w_i = 2/(1+\rho)$ if patient i contributes two teeth of a particular type

$= 1$ if patient i contributes one tooth of a particular type

It was necessary to estimate ρ, α and σ^2. This was achieved by using the computer package BATHSPLINE (Silverman and Watters, 1984). The estimate

of σ^2 used by the computer package is based on the definition in Silverman, (1985), namely,

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{m} \frac{w_i (\bar{Y}_{i\cdot} - \hat{f}_\alpha(x_i))^2}{[1 - \frac{1}{n} \mathrm{tr} A(\alpha)]}$$

where $\hat{f}_\alpha$ is the spline smoother and $\mathrm{tr}A\,(\alpha)$ is given in Silverman (1985) by

$$\mathrm{tr}A(\alpha) \doteq 2 + \sum_{i=3}^{m} \{1 + c_0 \alpha (i-1.5)^4\}^{-1}$$

From model (3.3.2)

$$Y_{i1} = f(x_i) + \tau_i + \eta_{i1}$$

$$Y_{i2} = f(x_i) + \tau_i + \eta_{i2}$$

so $Y_{i1} - Y_{i2} = \eta_{i1} - \eta_{i2}$. Hence $d_i = \left[Y_{i1} - Y_{i2}\right] \sim N(0, 2\sigma_2^2)$ since the η_{ik} are assumed independent.

Suppose there are m_1 patients contributing one tooth of a particular type and m_2 patients contributing two teeth of a particular type, then $m = m_1 + m_2$. An estimate of ρ was obtained by using an iterative process. Each iteration consisted of the following steps;

(i) Calculate an estimate of σ^2 using the BATHSPLINE package with weight w_i attached to knot i.

(ii) Calculate an estimate of σ_2^2 using

$$\hat{\sigma}_2^2 = \frac{1}{2(m_2 - 1)} \left\{ \sum_{i=1}^{m_2} d_i^2 - \frac{(\Sigma d_i)^2}{m_2} \right\}$$

(iii) Obtain an estimate of ρ using

$$\hat{\rho} = \frac{\hat{\sigma}^2 - \hat{\sigma}_2^2}{\hat{\sigma}^2}$$

(iv) Calculate the weights $w_i = \dfrac{2}{1+\hat{\rho}}$

Steps (i) - (iv) were repeated until successive estimates of ρ agreed to three decimal places. For the first iteration *only*, the w_i were all taken to be one.

In a previous analysis (Osborne, 1978) the linear model given by

$$Y_{ik} = \beta_0 + \beta_1 x_i + \varepsilon_{ik} \qquad i = 1,2,\ldots m \qquad k = 1 \text{ or } 2$$

was fitted to these same data sets. For the upper canine teeth *only*, it was not possible to obtain maximum likelihood estimates of $\beta_0, \beta_1, \sigma_1^2$ and σ_2^2 and the data obtained from the three patients each providing two upper canine teeth showed certain peculiarities. Since the estimate of σ_2^2 is based on these six observations it was decided to assume that the observations on the upper canine teeth were independent, i.e. $\rho = 0$. It would then be possible to compare the estimates of ξ from the non-parametric approach described in this chapter directly with the parametric estimates of ξ obtained as a result of assuming the above linear model (see Chapter 7).

Table 3.3 gives estimates of σ^2, α and ρ for all six types of teeth. Figs. 3.15 - 3.17 show the spline smoothers $\hat{f}_\alpha$ for the upper central, lower central, upper lateral and lower canine teeth. Fig. 3.16 shows that the upper lateral teeth data are completely unsuitable for calibration purposes as the spline smoother is almost horizontal over most of the calibration range. This was further confirmed when data on two new patients was received. The measurements of ITTM for these two new patients were both 2.7mm yet their ages were 47.0 years and 27.8 years. These two new observations have been plotted in Fig. 3.16. It was therefore decided not to analyse this data set further.

It was mentioned earlier in this section that the posterior distribution of γ is a truncated multivariate normal distribution , in particular,

$$\mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$$

The mode of this truncated multivariate normal distribution will be $\hat{\gamma}$ provided $\hat{\gamma} \in \Theta_{MON}$. However for the upper canine and lower lateral data sets, $\hat{\gamma} \notin \Theta_{MON}$. Figs. 3.18 and 3.19 show the spline estimates, $\hat{f}$, (dotted curves) for the upper canine and lower lateral data sets respectively, where $\hat{f}$ is given by

$$\hat{f}(x) = \sum_{i=1}^{n} \hat{\gamma}_i \beta_i(x)$$

and $\hat{\gamma} = (\hat{\gamma}_1 \hat{\gamma}_2 \hat{\gamma}_3 \ldots\ldots \hat{\gamma}_n)^T$.

It would therefore seem, that in these cases, we have to solve the following constrained optimisation problem: minimise

$$\sum_{i=1}^{n} w_i (f(x_i) - Y_i)^2 + \alpha \int_a^b [f''(x)]^2 dx$$

$$\text{subject to} \quad f \in W_2, \quad f \in \mathcal{F}_{MON} \quad \text{and} \quad \alpha > 0$$

where $\mathcal{F}_{MON}$ is the set of *monotonic* natural cubic smoothing splines with knots $\{x_i\}$ and W_2 is defined in Section 3.1.1.

Table 3.3 Details of the teeth calibration data

Type of tooth	Calibration range (yrs)	Total number of teeth (N)	Estimate of σ^2	Estimate of ρ	Estimate of α
Upper central	16.1 - 62.5	32	0.0142	0.502	513
Upper lateral	30.8 - 48.5	16	0.0128	0.980	18
Upper canine	20.3 - 62.5	15	0.0096	Assumed zero	214
Lower central	26.6 - 62.5	28	0.0252	0.606	183
Lower lateral	27.8 - 62.5	28	0.0225	0.634	176
Lower canine	22.8 - 65.6	24	0.0069	0.450	298

Fig.3.15: Upper and lower central teeth

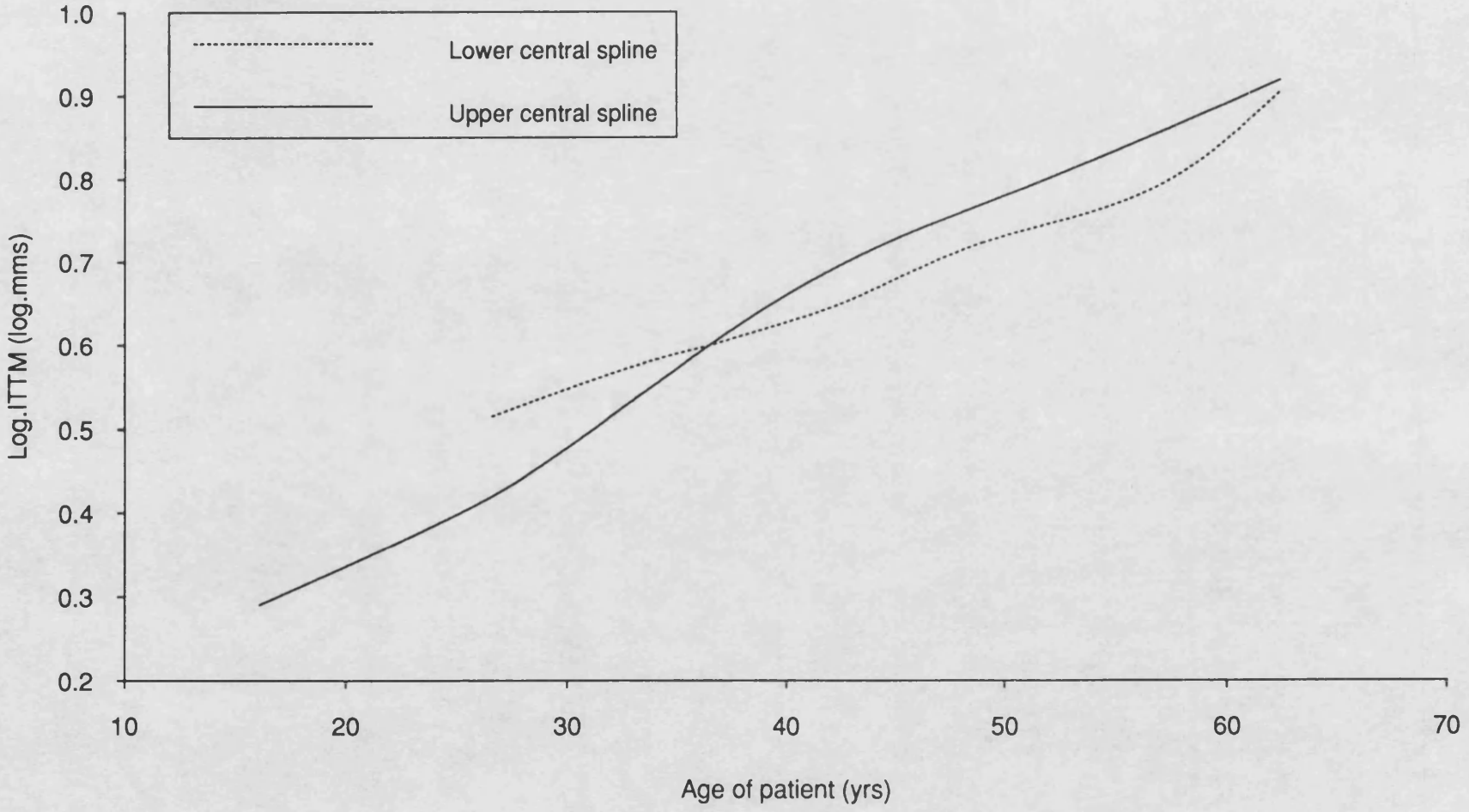


Fig.3.16: Upper lateral teeth

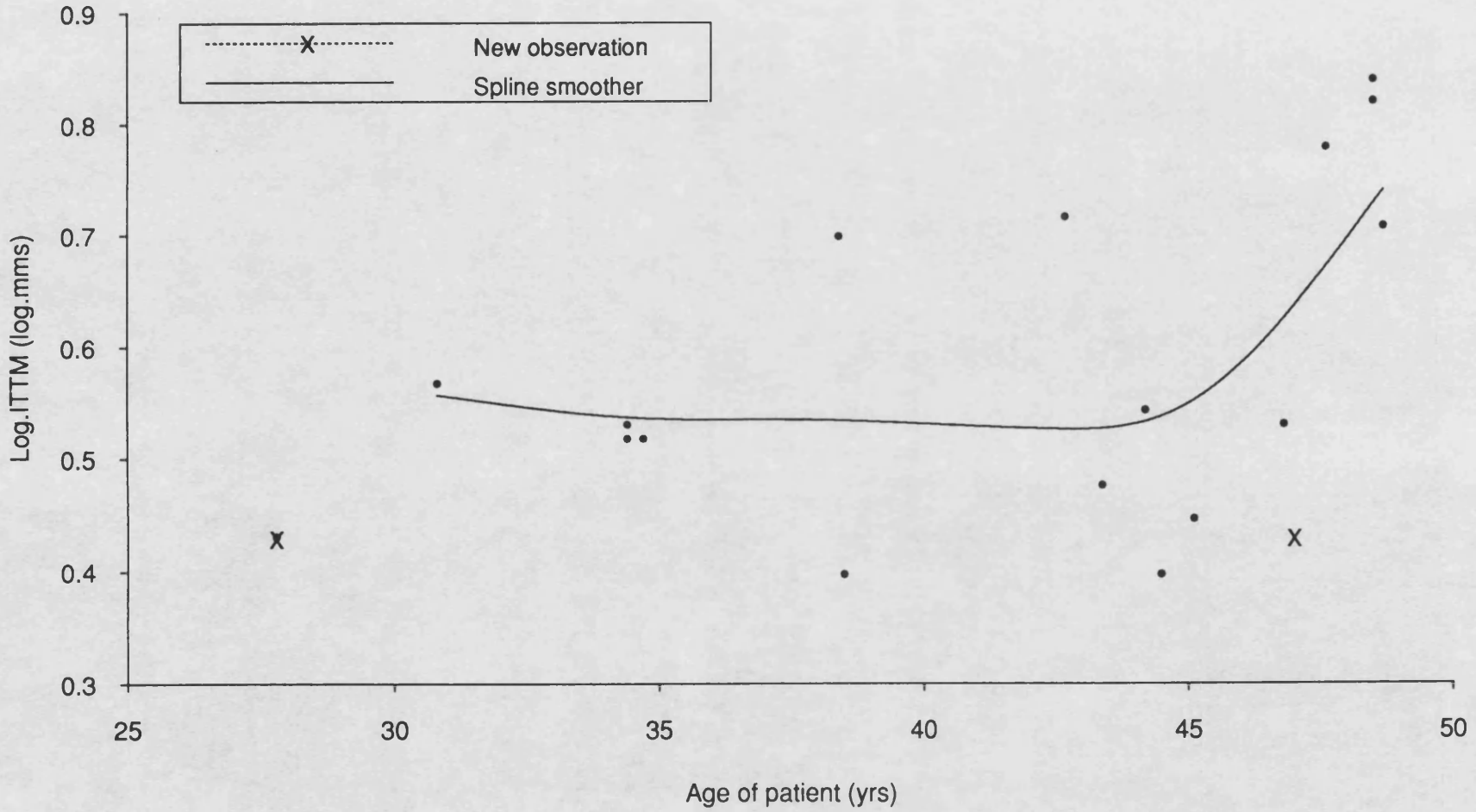


Fig.3.17: Lower canine teeth

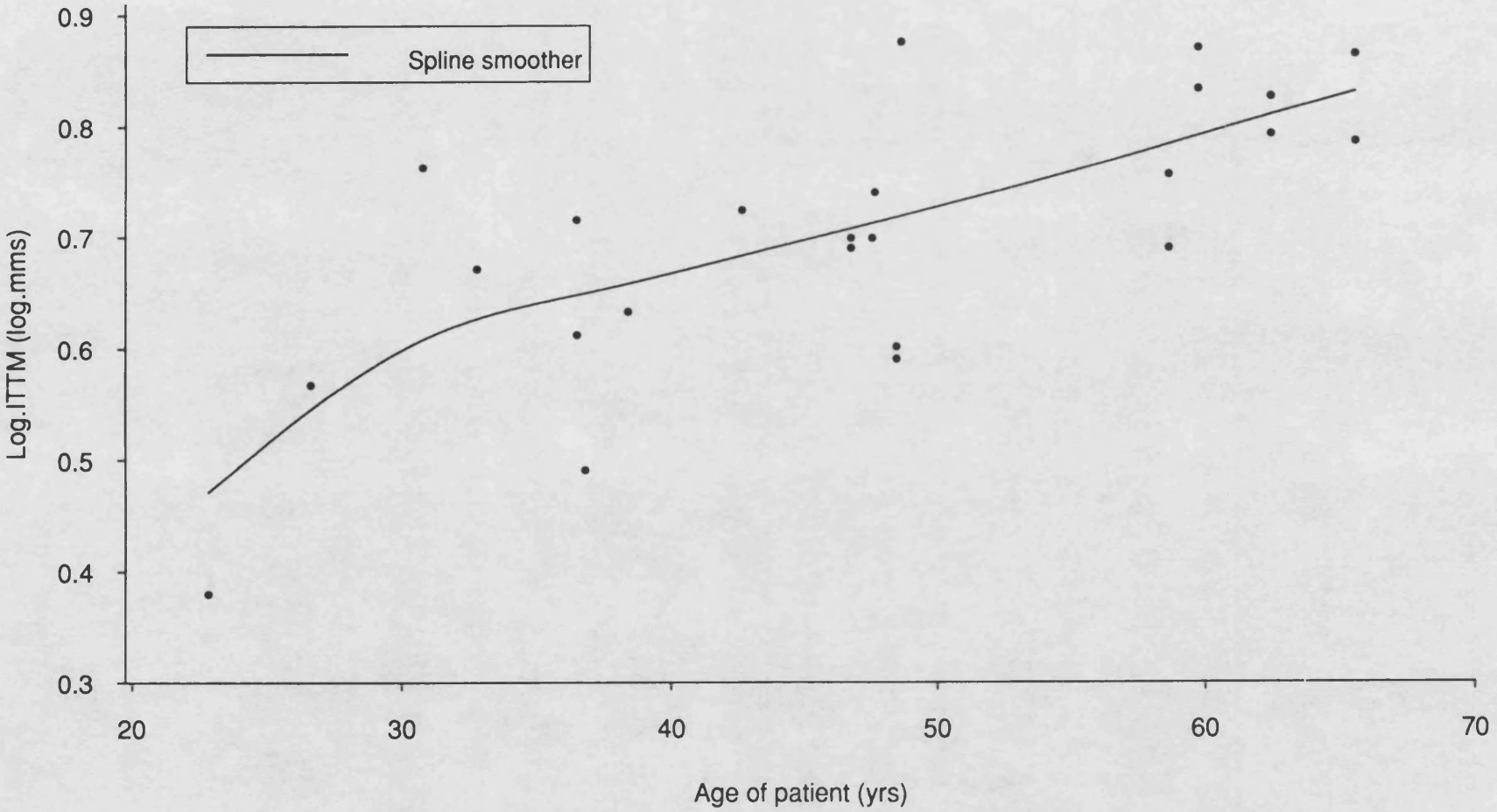


Fig.3.18: Upper canine teeth

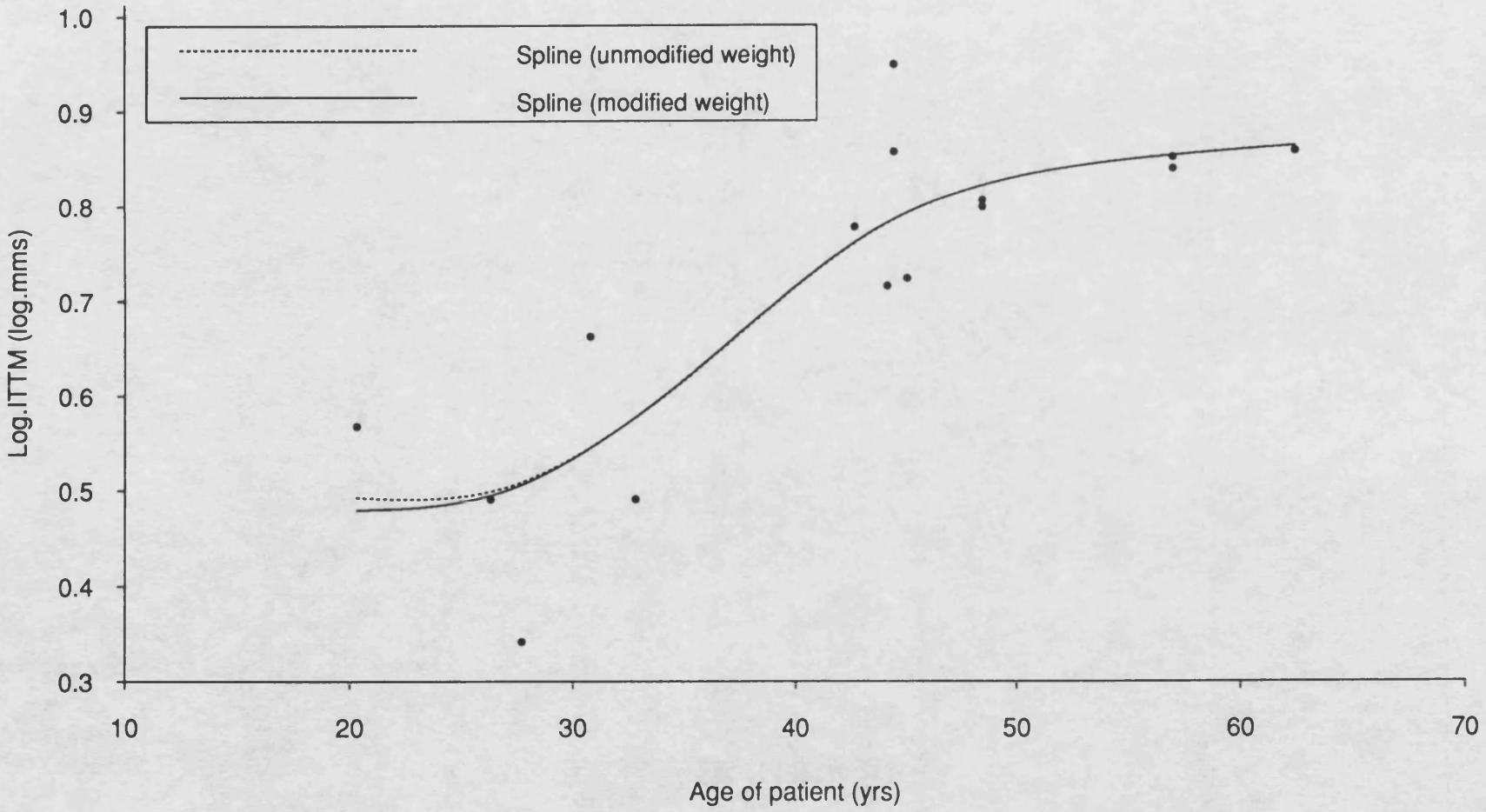
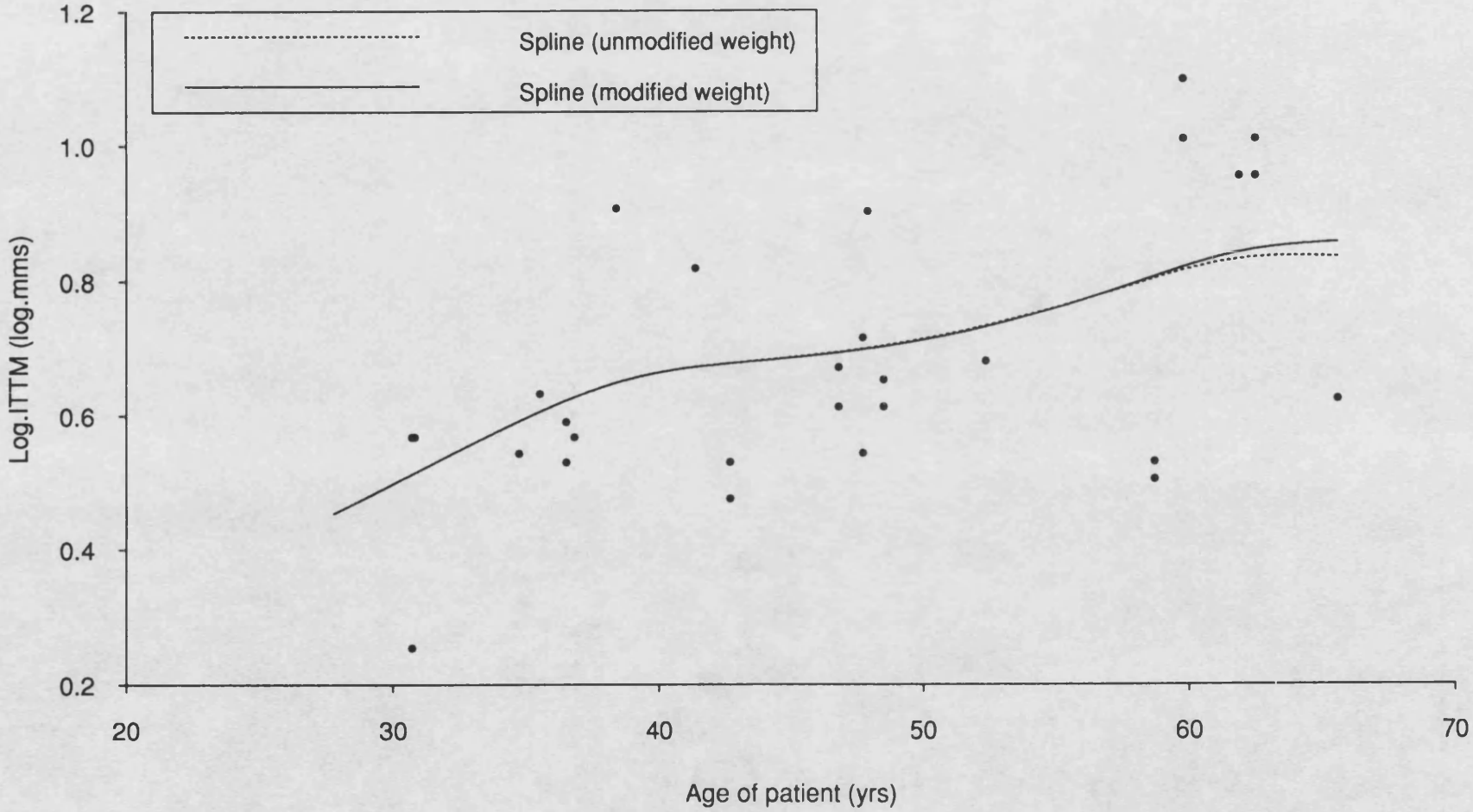


Fig.3.19: Lower lateral teeth



This constrained optimisation problem is rather intractable so it was decided to adopt an *ad hoc* procedure which was very easy to implement and which overcame the difficulty of $\hat{\gamma} \notin \Theta_{MON}$.

It will be noted from Figs. 3.18 and 3.19 that the lack of monotonicity is confined to the edges of the calibration range, the lower end of the range for the upper canine data (see Fig. 3.18) and the upper end of the range for the lower lateral data (see Fig. 3.19). Extreme data values exert excessive leverage on splines. The *ad hoc* procedure thus consisted of altering the leverage of these extreme values by adjusting the weights attached to extreme points. In particular, for the upper canine data set, the weight attached to the point (20.3, log 3.7) was altered from 1 to 0.80 and for the lower lateral data set, the weight attached to the point (65.6, log 4.2) was adjusted from 1 to 0.80. Then, as for the other data sets,

$$\sum_{i=1}^{n} w_i^* (f(x_i)-Y_i)^2 + \alpha \int_a^b [f''(x)]^2 dx$$

was minimised subject to $f \in W_2$, and $\alpha > 0$

using the BATHSPLINE program (Silverman and Watters, 1984). Here w_i^* denotes the adjusted weight attached to observation (x_i, Y_i). The *ad hoc* procedure only involved *one* of these weights being adjusted, in particular w_1 for the upper canine data set and w_n for the lower lateral data set. The resulting spline estimates, $\hat{f}_\alpha$, for the upper canine and lower lateral data sets are shown in Figs. 3.18 and 3.19 respectively.

The ultimate objective of this data analysis was to predict the age of a forensic victim given observations $Y_1^*, Y_2^*, \ldots Y_m^*$ on certain teeth belonging to the victim and the data from the calibration experiment which relates to patients not victims. Let $X =$ age of a human being and $Y =$ logarithm of the length of transparent root dentine for a particular tooth belonging to that human being. Let X and Y possess a bivariate distribution with density $p_p(x,y)$ if the human being is a patient and density function $p_v(x,y)$ if the human being is a victim whose body is to be identified by forensic odontological methods. The question arises as to what common features these bivariate distributions possess. Does one assume that the two bivariate distributions are identical or does one assume that the conditional distributions $p_p(y|x)$ and $p_v(y|x)$ are identical? It was decided that the latter assumption was more realistic. The objective of the data analysis, therefore, was to arrive at the conditional distribution $p_v(\xi|Y^*)$ where ξ is the unknown age of a victim and Y^* is a reading of Y available from the victim's teeth. The model for the prediction stage of the calibration process is given by;

$$Y_j^* = f(\xi) + \varepsilon_j^* \qquad j=1,2,\ldots,m$$

where f is a natural cubic smoothing spline with knots $\{x_i\}$ $i=1,2,\ldots,n$ and the ε_j^* are assumed to be i.i.d. normal random variables with mean zero and variance σ^2. However, no data on victims were available so the question arose as to how to assess this non-parametric method of estimation. This was achieved by obtaining information on 31 new patients by making measurements of Y, namely $Y_1', Y_2', \ldots, Y_m'$ on the teeth of these new patients and obtaining estimates of the age (ξ) of these new patients. The estimated ages were then compared with the actual ages of the patients. The model for the prediction stage is therefore given by;

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,\ldots,m$$

where it is assumed that the ε_j' are i.i.d. normal random variables with mean zero and variance σ^2. The largest value of m was 4 but for most new patients, $m = 1$ or 2 for any particular type of tooth. Posterior distributions of ξ were obtained as described in the previous section.

Table 3.4 gives the mean absolute error for each of the five types of teeth considered, together with interval estimates for ξ. (Estimates were calculated to the nearest tenth of a year). A probability level of 80% was acceptable to the forensic odontologist. An interval for ξ was deemed to be *successful* if the age of the new patient lay inside the estimated interval. These results compare favourably with those obtained by using parametric approaches where it is assumed that f is a straight line (see Chapter 7). What is pleasing about the method is that it works reasonably well even when the number of observations (N) in the calibration experiment is quite small as is the case for the upper canine data where $N = 15$.

For all types of teeth, the scatter in the calibration data is considerable. As there is variation in the size of teeth, it might be argued that the proportion of transparent root dentine in relation to the total amount of root dentine ought to be measured rather than the absolute amount of transparent root dentine. Such measurements were made on the upper canine teeth and the analysis of these results is made in Chapter 6.

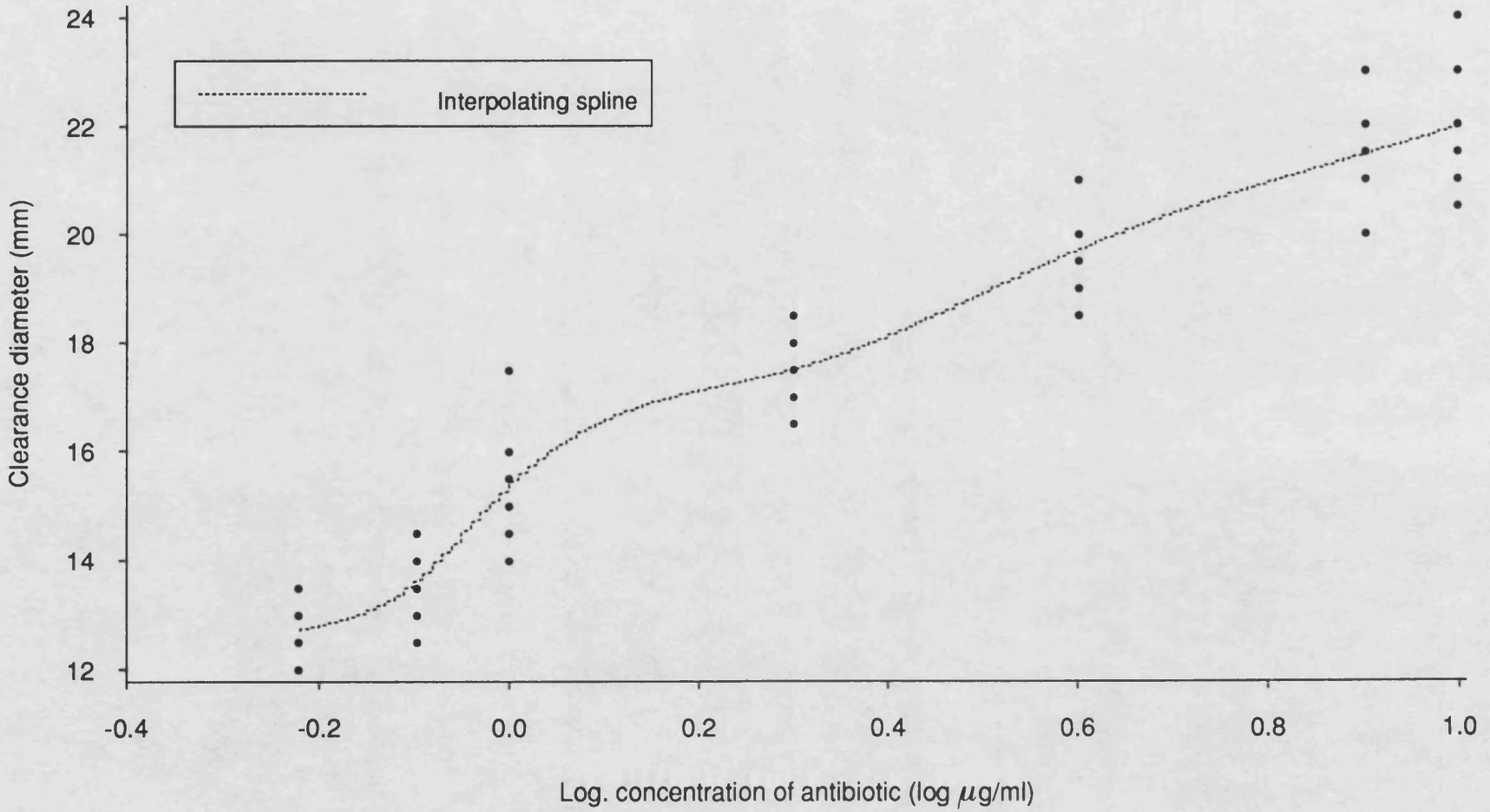
3.3.3 Antibiotic assay data

The calibration data considered in this section consist of measurements of the clearance diameter (D) of 67 circles which were formed when droplets of antibiotic of a known concentration were added to an infected medium. Let C = log (concentration of antibiotic). Fig. 3.20 shows a scatter diagram of the calibration data.

Table 3.4 Mean absolute error and interval estimates for ξ

Type of tooth	Upper central	Lower central	Upper canine	Lower lateral	Lower canine
Mean absolute error (yrs)	7.5	5.7	4.6	5.6	10.1
Maximum width of interval estimates (yrs)	21.8	26.1	19.1	24.8	28.1
Minimum width of interval estimates (yrs)	13.0	9.5	11.2	12.4	11.7
Mean width of interval estimates (yrs)	17.0	15.2	15.3	20.6	20.0
No. of successful estimates	10	3	5	5	9
Total no. of new teeth	15	3	7	6	13

Fig.3.20: Antibiotic assay data



From Fig. 3.20, it can be seen that there are seven chosen levels of C in the calibration experiment. Full details of this data set are given in Appendix 1.

The first stage to obtaining the posterior distribution of ξ, where ξ is the *unknown* log concentration of antibiotic in the blood of a patient, is to simulate from the posterior distribution of f. So consider the following model for this calibration experiment

$$D_{ij} = \mu_i + \varepsilon_{ij} \qquad \begin{matrix} j=1,2,\ldots,n_i \\ i=1,2,3,\ldots,7 \end{matrix} \tag{3.3.3}$$

where D_{ij} is the jth clearance diameter at level i of C. Suppose for simplicity that $j=1,2,\ldots,m$ i.e. an equal number of replications at each level i of C. Assume also that the ε_{ij} are i.i.d normal random variables with mean zero and variance σ^2. Each $\bar{D}_{i\cdot}$ $(i=1,2,\ldots,7)$ is normally distributed with mean μ_i and variance σ^2/m where

$$\bar{D}_{i\cdot} = \sum_{j=1}^{m} D_{ij}/m$$

Let the $i\times 1$ vectors $\bar{\mathbf{D}}$ and θ be defined as follows:

$$\bar{\mathbf{D}}^{\mathrm{T}} = (\bar{D}_{1\cdot}\ \bar{D}_{2\cdot}\ \bar{D}_{3\cdot}\ \ldots\ldots\ \bar{D}_{7\cdot})$$

$$\theta^{\mathrm{T}} = (\mu_1\ \mu_2\ \mu_3 \ldots\ldots\ \mu_7)$$

Then $\bar{\mathbf{D}}$ has a multivariate normal distribution with mean θ and variance matrix $\frac{\sigma^2}{m} I_7$. Consider a locally uniform prior for θ given by

$$\pi(\theta \mid \sigma) \propto k$$

where k is a constant, which is then truncated to remove those values of θ which do not produce monotonicity. Then the posterior distribution of θ is a truncated multivariate normal distribution. Using the notation of Section 3.2.2, the posterior distribution of θ is given by

$$\mathcal{N}_{MON}(\bar{\mathbf{D}}, \frac{\sigma^2}{m} I_7)$$

To obtain posterior realisations of f, the steps are:

(a) Obtain posterior realisations of θ, θ^*, where $\theta^{*\mathrm{T}} = (\theta_1^*\ \theta_2^* \ldots\ldots \theta_7^*)$ and $\theta_i^* = \bar{D}_{i\cdot} + z\sigma/\sqrt{m}$ where z is a standard normal random variable.

(b) For each realisation of θ, find a natural cubic spline which interpolates the points (c_i, θ_i^*) $i=1,2,3,\ldots,7$.

(c) Reject any realisations of θ which produce a spline which is not monotonic increasing.

If one compares model (3.3.3) with the model (3.2.1a), namely

$$D_{ij} = f(c_i) + \varepsilon_{ij} \qquad i=1,2,\ldots,7$$

where f is a monotonic natural cubic spline with knots $\{c_i\}$, one can see that

$$f(c_i) = \mu_i \qquad i=1,2,\ldots,7. \tag{3.3.4}$$

Result (3.3.4) can also be derived by putting $\alpha = 0$ in equations (3.3.1) with $\mathbf{Y} = \overline{\mathbf{D}}$ and $W = mI_7$ so $S^{-1} = \frac{\sigma^2}{m}(B^{\mathrm{T}}B)^{-1}$ and $\hat{\gamma} = B^{-1}\overline{\mathbf{D}}$.

If one defines $\mathbf{f}^{\mathrm{T}}$ as the 1×7 row vector

$$(f(c_1)\ f(c_2)\ f(c_3)\ \ldots\ f(c_7))$$

since $\mathbf{f} = B\gamma$ from equation (3.1.12), it follows that $\mathbf{f}$ has a truncated multivariate normal distribution namely,

$$\mathcal{N}_{MON}(B\hat{\gamma}, BS^{-1}B^{\mathrm{T}}) \equiv \mathcal{N}_{MON}(\overline{\mathbf{D}}, \frac{\sigma^2}{m}I_7)$$

Hence $\mathbf{f} = \boldsymbol{\theta}$.

Assuming the untruncated prior distribution of $\boldsymbol{\theta}$ was locally uniform is equivalent to assuming that the untruncated prior distribution of γ is locally uniform instead of a multivariate normal distribution.

It was therefore not necessary to create a new computer program to perform steps (a) to (c). The program used for generating posterior realisations of f for the previously considered data sets was also used for this data set, only taking the smoothing parameter α equal to zero. Because α is zero, f is a natural cubic *interpolating* spline, interpolating the points (c_i, θ_i^*) $i=1,2,3,\ldots,7$ so that

$$\tilde{f}(c_i) = \theta_i^* \qquad i=1,2,3,\ldots,7$$

where θ_i^* is a posterior realisation of μ_i and $\tilde{f}$ is a posterior realisation of f.

The parameter σ^2 was estimated by using the pooled *within* sums of squares, i.e.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{7}\sum_{j=1}^{n_i}(D_{ij}-\overline{D}_{i\cdot})^2}{\sum_{i=1}^{7}(n_i-1)}$$

Here $n_1 = 7$ and $n_i = 10$, $i=2,3,4,5,6,7$. Fig. 3.20 shows the interpolating spline estimate, $\hat{f}_I$, for this data set.

In order to test out the performance of the method with this data set, single measurements of clearance diameter (D) were made on 60 new circles formed in the same infected medium. Point and interval estimates of ξ were obtained from the posterior distribution of ξ. Table 3.5 gives the mean absolute error and details of the 60 interval estimates (the probability level is 90% for the interval estimates).

3.4 U-STATISTICS

In Section 3.2.2 of this chapter the posterior distribution of ξ was obtained by generating m posterior realisations of η, η_v, for each of n posterior realisations of f, $\tilde{f}_s$, giving mn posterior realisations of ξ, X_{vs} where

$$X_{vs} = \tilde{f}_s^{-1}(\eta_v) \qquad \begin{matrix} v = 1,2,\ldots,m \\ s = 1,2,\ldots,n. \end{matrix}$$

A *different* set of values η_v were generated for each posterior realisation of f, $\tilde{f}_s$. The choice of $m = 50$ and $n = 200$ was rather arbitrary. The question therefore arises as to whether it is possible to choose m and n in some *optimal* way given a total time, C cpu, available on the computer. This section will consider the problem of choosing m and n optimally.

Let $h(Y_v, Z_s)$ be the point at which the spline Z_s crosses the value Y_v and let

$$h(Y_v, Z_s) = X_{vs} \qquad \begin{matrix} v = 1,2,\ldots,m \\ s = 1,2,\ldots n \end{matrix}$$

Let $Y_1 Y_2,\ldots,Y_m$ and $Z_1 Z_2,\ldots,Z_n$ be independent, the Y's identically distributed with distribution F and the Z's identically distributed with distribution G. Let

$$W = \sum_{v=1}^{m} \sum_{s=1}^{n} h(Y_v, Z_s) = \sum_{v} \sum_{s} X_{vs} \qquad \text{and let } U = \frac{1}{mn} W.$$

Then U is a U-statistic (Lehmann, 1975) and

$$\begin{aligned} \text{Var}(W) = {} & mn\,\text{Var}(X_{vs}) + mn(m-1)\,\text{Cov}\,(X_{vs}, X_{ws}) \\ & + mn(n-1)\,\text{Cov}\,(X_{vs}, X_{vt}) \qquad \begin{matrix} v \neq w \\ s \neq t \end{matrix} \end{aligned}$$

Consider Cov (X_{vs}, X_{ws}). This covariance term arises from the *different* Y values (Y_v and Y_w) on the same spline Z_s. The values $X_{1s}, X_{2s}, X_{3s}, \ldots X_{ms}$ are not independent because the points (X_{1s}, Y_1) $(X_{2s}, Y_2),\ldots,(X_{ms}, Y_m)$ are constrained to lie on the spline Z_s. Hence Cov $(X_{vs}, X_{ws}) \neq 0$.

Table 3.5 Point and interval estimates for ξ

Mean absolute error	0.069
Maximum width of interval estimates	0.45
Minimum width of interval estimates	0.14
Mean width of interval estimates	0.30
No. of successful estimates	52
Total no. of estimates	60
Calibration range	-0.22 to 1.00

Consider the covariance term Cov (X_{vs}, X_{vt}). This term arises from the *same* value Y_v and *different* splines Z_s and Z_t. In the simulation outlined in Section 3.2.2 *different* values Y_v (v=1,2,...,m) were chosen for each spline Z_s (s=1,2,...,n) so the observations X_{vs}, X_{vt} are independent. Hence Cov (X_{vs}, X_{vt}) = 0 and

$$\text{Var(W)} = mn\,\text{Var}(X_{vs}) + mn(m-1)\,\text{Cov}\,(X_{vs}, X_{ws})$$

$$\text{Var}(U) = \frac{\text{Var}(X_{vs})}{mn} + \frac{(m-1)}{mn}\,\text{Cov}\,(X_{vs}, X_{ws}) \qquad v \neq w \qquad (3.4.1)$$

The above result can also be obtained using a different approach and it is useful to consider this approach too.

Let X_{ij} be the posterior realisation of ξ given by

$$X_{ij} = \tilde{f}_i^{-1}(\eta_j)$$

Consider the following model

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \begin{matrix} i=1,2,\ldots,n \\ j=1,2,\ldots,m \end{matrix} \qquad (3.4.2)$$

where it is assumed that the τ_i are i.i.d. random variables with mean zero and variance σ_τ^2. Assume the ε_{ij} are i.i.d. random variables with mean zero and variance σ_ε^2 and that the ε_{ij} are independent of the τ_i . Assume without loss of generality that μ = 0. Here i denotes the number of the spline (i=1,2,...,n) and j denotes the number of the η value (j=1,2,...,m), so τ_i relates to the variation in X values arising from different splines whereas ε_{ij} relates to the variation in X values arising from different η values on the ith spline. It follows from model (3.4.2) and its assumptions, that X_{ij} is independent of $X_{i'j'}$ for $i \neq i'$.

Let $X_{i\cdot} = \sum_{j=1}^{m} X_{ij}$ and $\varepsilon_{i\cdot} = \sum_{j=1}^{m} \varepsilon_{ij}$ then $X_{i\cdot} = m\tau_i + \varepsilon_{i\cdot}$.

$$E(X_{i\cdot}^2) = m^2\sigma_\tau^2 + m\sigma_\varepsilon^2$$

Also $$\text{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\right] = \text{Var}\left[\sum_{i=1}^{n} X_{i\cdot}\right] = \sum_{i=1}^{n} \text{Var}\,(X_{i\cdot})$$

since $X_{i\cdot}$ is independent of $X_{i'\cdot}$ for $i \neq i'$. Hence

$$\text{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\right] = \sum_{i=1}^{n} E(X_{i\cdot}^2) \quad \text{since} \quad E(X_{i\cdot}) = 0$$

It follows that

$$\text{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\right] = nm^2\sigma_\tau^2 + mn\sigma_\varepsilon^2$$

$$\text{Var}\left[\frac{\sum_i \sum_j X_{ij}}{mn}\right] = \frac{\sigma_\tau^2}{n} + \frac{\sigma_\varepsilon^2}{mn} \tag{3.4.3}$$

$$\begin{aligned}\text{Cov}\,(X_{ij}, X_{ij'}) &= \sigma_\tau^2 + \sigma_\varepsilon^2 \qquad j = j' \\ &= \sigma_\tau^2 \qquad\qquad j \neq j'\end{aligned} \tag{3.4.4}$$

From equation (3.4.3)

$$\text{Var}\left[\frac{\sum\sum X_{ij}}{mn}\right] = \frac{\sigma_\tau^2}{n} + \frac{1}{mn}\left[\sigma_\tau^2+\sigma_\varepsilon^2-\sigma_\tau^2\right]$$

$$= \frac{(\sigma_\tau^2+\sigma_\varepsilon^2)}{mn} + \frac{(m-1)}{mn}\sigma_\tau^2$$

$$= \frac{\text{Var}(X_{ij})}{mn} + \frac{(m-1)}{mn}\text{Cov}\,(X_{ij}, X_{ij'}) \qquad j \neq j'$$

$$\text{So} \qquad \text{Var}\,(U) = \frac{\text{Var}(X_{ij})}{mn} + \frac{(m-1)}{mn}\text{Cov}\,(X_{ij}, X_{ij'}) \quad j \neq j'$$

This result is identical to equation (3.4.1).

Suppose one takes $n = 200$, $m = 1$, then $\text{Cov}\,(X_{ij}, X_{ij'}) = 0$, so in this case

$$\text{Var}\,(U) = \frac{\text{Var}(X_{ij})}{200} = \frac{\sigma_\tau^2+\sigma_\varepsilon^2}{200} \tag{3.4.5}$$

Suppose one takes $n = 10000$, $m = 1$ then again $\text{Cov}\,(X_{ij}, X_{ij'}) = 0$ and

$$\text{Var}\,(U) = \frac{\sigma_\tau^2+\sigma_\varepsilon^2}{10{,}000}. \tag{3.4.6}$$

In the simulation experiment outlined in this chapter (Section 3.2.2) the values of m and n were taken to be 50 and 200 respectively. Substituting for m and n in equation (3.4.3) gives

$$\text{Var}\,(U) = \frac{\sigma_\tau^2}{200} + \frac{\sigma_\varepsilon^2}{10{,}000} \tag{3.4.7}$$

If one compares equations (3.4.5), (3.4.6) and (3.4.7) one can see that the simulation experiment provides more information on ξ then 200 independent posterior realisations of ξ would, but not as much information on ξ as 10,000 independent posterior realisations of ξ would provide.

We will now consider the problem of choosing m and n to minimise $\text{Var}(U)$ where $\text{Var}(U)$ is given by equation (3.4.1). Theorem 3.1 below shows a very

interesting result, namely that it is *only* necessary to choose m optimally. Having chosen m optimally then one just generates as many splines (n) as one requires or as computer time dictates.

Let $V_1 = \text{Var}(X_{vs})$ and $V_2 = \text{Cov}\,(X_{vs}, X_{ws})$ in equation (3.4.1), then

$$\text{Var}(U) = \frac{V_1}{mn} + \frac{(m-1)V_2}{mn}$$

Theorem 3.1.

Let t_s be the time in cpu to construct one monotonic spline and let t_v be the time in cpu to construct one posterior realisation of ξ given the spline.

The optimal value of m is given by

$$m_{\text{opt}} = \left[\frac{(V_1 - V_2)t_s}{V_2 t_v} \right]^{\frac{1}{2}} \tag{3.4.8}$$

Proof. Let C cpu be the total time available on the computer. We have a constrained minimisation problem here. We wish to minimise

$$\text{Var}\,(U) = \frac{1}{n}\left[\frac{V_1}{m} + \frac{(m-1)V_2}{m} \right] \tag{3.4.9}$$

subject to $$n t_s + mnt_v = C \tag{3.4.10}$$

From equation (3.4.10)

$$n = \frac{C}{(t_s + mt_v)}$$

Substituting into (3.4.9) gives

$$\text{Var}\,(U) = \frac{(t_s + m t_v)}{C}\left[\frac{V_1}{m} + \frac{(m-1)V_2}{m} \right]$$

$$= \frac{(t_s + m t_v)}{C}\left[\frac{(V_1 - V_2)}{m} + V_2 \right]$$

Differentiating the above with respect to m and equating to zero gives

$$\frac{d}{dm}\,[\text{Var}(U)] = \frac{1}{C}\left[V_2 t_v - \frac{(V_1 - V_2)t_s}{m^2} \right] = 0$$

So $m^2 V_2 t_v = (V_1 - V_2)t_s$ i.e.

$$m = \left[\left(\frac{V_1 - V_2}{V_2} \right) \frac{t_s}{t_v} \right]^{\frac{1}{2}} \tag{3.4.11}$$

Differentiating again gives

$$\frac{d^2}{dm^2}\,[\mathrm{Var}(U)] = \frac{1}{C}\left[\frac{2(V_1-V_2)t_s}{m^3}\right] = \frac{2V_2\,t_v}{Cm}$$

The value of m given in (3.4.11) minimises $\mathrm{Var}(U)$. Hence the optimal value of m is given by

$$m_{\mathrm{opt}} = \left[\frac{(V_1-V_2)t_s}{V_2 t_v}\right]^{\frac{1}{2}}$$

3.4.1 Application of theorem to simulated data sets

The theorem was applied to the Gompertz and Weibull data sets. An estimate of V_1 was made by using the variance of the posterior distribution of ξ. This was readily available. The question arose as to how to estimate V_2. It will be recalled that $V_2 = \mathrm{Cov}\,(X_{vs}, X_{ws})$ $\mathrm{v} \neq \mathrm{w}$. Using the notation of model (3.4.2)

$$V_2 = \mathrm{Cov}\,(X_{ij}, X_{ik}) \qquad j \neq k$$

where i denotes the ith spline and j,k denote the jth and kth η values. So this covariance term is related to the variation in X values which arises from *different* η values on the *same* spline.

Model (3.4.2) is a one-way random effects model. The ANOVA table for this model is given by

Source	SS	d.f.	MS	E(MS)
Between splines	SS_B	$(n-1)$	$SS_B/(n-1)$	$m\sigma_\tau^2 + \sigma_\varepsilon^2$
Within splines	SS_e	$n(m-1)$	$SS_e/n(m-1)$	σ_ε^2
Total	SS_T	$(mn-1)$		

where

$$SS_B = m\sum_{i=1}^{n}(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2 \qquad \bar{X}_{i\cdot} = \frac{\sum_j X_{ij}}{m}$$

$$SS_e = \sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij}-\bar{X}_{i\cdot})^2 \qquad \bar{X}_{\cdot\cdot} = \frac{\sum_i\sum_j X_{ij}}{mn}$$

$$SS_T = \sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij}-\bar{X}_{\cdot\cdot})^2$$

Also $SS_B + SS_e = SS_T$ so $SS_B = SS_T - SS_e$.

Then according to Scheffé (1959)

$$\hat{\sigma}_\tau^2 = \frac{1}{m}\left[\frac{SS_B}{(n-1)} - \frac{SS_e}{n(m-1)}\right] \qquad (3.4.12)$$

so equation (3.4.12) was used to obtain an estimate of V_2.

As the time to generate a spline and a posterior realisation of ξ was so small t_v and t_s were estimated by calculating the time to generate 3000 splines and 3000 values of ξ and dividing this time by 3000. Table 3.6 gives the values of t_s, t_v, the maximum value of m_{opt}, the minimum value of m_{opt} and the mean value of m_{opt} for the Gompertz and Weibull data sets. It will be noted that though, as expected, the value of t_v is the same for both data sets, the value of t_s is different being much larger for the Gompertz data set. The time t_s is the time to generate a *monotonic* spline. It is greater for the Gompertz data set because a higher percentage of generated splines have to be rejected (because of their lack of monotonicity) compared to the Weibull data set. If one compares Figs. 3.7 and 3.8, one will see that the underlying model for the Gompertz data set has a flatter region at the lower end of the calibration range than the underlying model for the Weibull data set.

It should be noted that although the optimal value of m was not used to generate the posterior distributions of ξ for any of the data sets considered in this chapter, the resulting estimates of ξ were still precise. However, equation (3.4.8) is used in Chapter 6 to generate the optimal m.

3.5 CONCLUDING REMARKS

The method considered in this chapter viewed ξ as a non-linear functional of the true calibration curve f. Simulation was used to generate the posterior distribution of ξ. The method involved

(a) generating posterior realisations of f, where f is a monotonic cubic smoothing spline with knots at the data points $\{x_i\}$ $i=1,2,3....n$,

(b) generating posterior realisations of η,

(c) calculating $\tilde{f}_s^{-1}(\eta_v)$, where $\tilde{f}_s$ denotes a posterior realisation of f and η_v denotes a posterior realisation of η.

The posterior distribution of f is a truncated multivariate normal distribution whereas the posterior distribution of η is a truncated normal distribution.

Table 3.6 Optimal values of m for the Weibull and Gompertz data sets

	Weibull	Gompertz
Time to generate one spline (cpu) t_s	0.025	0.058
Time to generate one posterior realisation of ξ given the spline (cpu) t_v	0.002	0.002
Maximum value of m_{opt}	17	26
Minimum value of m_{opt}	12	19
Mean value of m_{opt}	14.15	23.0

A simple formula enables one to quickly calculate the *optimal* number of values, η_v, to be generated from the posterior distribution of η, for a given posterior realisation of f.

The method showed itself to be very versatile, working well on all the data sets. The posterior median was used as a point estimate for ξ and symmetrical interval estimates for ξ were constructed from the posterior distributions of ξ. Both the point and interval estimates for ξ were extremely good for all the data sets. The advantages of this method were that it was extremely easy to implement and the amount of computer time required to generate a posterior distribution of ξ was extremely small. There seemed to be no obvious disadvantages.

4. AN APPROACH USING SIMULATION AND AN EXPLICIT PRIOR

4.1 INTRODUCTION

In Chapter 3, the following model for the calibration experiment was used

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,3.....n \tag{4.1.1}$$

where f is a monotonic natural cubic spline with knots $\{x_i\}$ and the ε_i are assumed to be i.i.d normal random variables with mean zero and variance σ^2/w_i. For the prediction stage the model was given by

$$\begin{aligned} Y_j' &= f(\xi) + \varepsilon_j' \\ &= \eta + \varepsilon_j' \qquad j=1,2....m \end{aligned} \tag{4.1.2}$$

where again f is a monotonic natural cubic spline and the ε_j' are i.i.d normal random variables with mean zero and variance σ^2 independent of the ε_i.

There were three steps to generating the posterior distribution of ξ. The first step was to generate posterior realisations of $f, \tilde{f}_s$. The second step was to generate realisations from the posterior distribution of η, η_v. To this end, a uniform prior for η was assumed, namely

$$\begin{aligned} p(\eta) &= c \qquad \tilde{f}_s(x_1) \le \eta \le \tilde{f}_s(x_n) \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

where c is a constant. The likelihood associated with the m observations Y_j' is given by

$$\mathcal{L}(\eta) = \left[\frac{m}{2\pi\sigma^2}\right]^{\frac{1}{2}} \exp\left[\frac{-m}{2\sigma^2}(\eta-\bar{Y}')^2\right] \tag{4.1.3}$$

where $\bar{Y}'$ = mean of the m observations $Y_1', Y_2', \ldots, Y_m'$. Hence the posterior density of η is given by

$$p(\eta|\mathbf{Y}') = C\left[\frac{m}{2\pi\sigma^2}\right]^{\frac{1}{2}} \exp\left[\frac{-m}{2\sigma^2}(\eta-\bar{Y}')^2\right] \qquad \tilde{f}_s(x_1) \le \eta \le \tilde{f}_s(x_n)$$

$$= 0 \qquad \text{otherwise}$$

The resulting posterior distribution of η is a truncated normal distribution. Since $\xi = f^{-1}(\eta)$, the final step of the process to generating the posterior distribution of ξ was to calculate $f_s^{-1}(\eta_v)$, for as many values of s and v as were required.

The criticism which could be levelled at this approach is that it is difficult to incorporate prior information about ξ. Lawless (1981) criticised Hunter and Lamboy's (1981a) parametric approach to linear calibration on the same grounds. Hunter and Lamboy(1981a) assumed a non-informative prior (a locally uniform prior) for η which implied a prior for ξ which was also non-informative since

$$\xi = \frac{\eta - \beta_0}{\beta_1}$$

where β_0 and β_1 are the intercept and slope parameters of the true calibration line. In contrast, Hoadley (1970) used an approach via predictive densities which enabled prior information on ξ, in the form of a prior distribution for ξ, to be used quite easily.

This chapter outlines an approach which still uses non-linear functional ideas but which enables prior information on ξ, in the form of a prior distribution, to be directly utilised in obtaining the posterior distribution of ξ. The models are the same as those used in Chapter 3, namely equations (4.1.1) and (4.1.2) given above. The notation is also the same as that of Chapter 3. A non-parametric approach which is based on predictive densities and which is akin to the parametric approach of Hoadley (1970) is considered in the next chapter, Chapter 5.

4.2 REJECTION SAMPLING

Suppose it is wished to simulate from the p.d.f. $f(x)$. Let $h(x)$ be a p.d.f. with the same range as $f(x)$ but from which it is easier to simulate. The basic idea of rejection sampling (Morgan, 1984) is to envelop $f(x)$ by the curve $w(x) = kh(x)$ where k is a constant ($k > 1$). A uniform scatter of points (X,Z) is generated under $w(x)$ by simulating X from the distribution with p.d.f. $h(x)$ and taking the conditional density of Z given $X = x$ to be a uniform density on $(0, w(x))$ i.e. $U(0,w(x))$. Points which lie below $w(x)$ but above $f(x)$ are rejected. Since $w(x) \geq f(x)$ for all x, k is defined as follows:

$$k = \sup_x \frac{f(x)}{h(x)}.$$

The following algorithm for rejection sampling is thus obtained:

(i) Simulate $X = x^*$ from the probability density $h(x)$;

(ii) Simulate $Z = U\, h(x^*) \sup_x \frac{f(x)}{h(x)}$ where U is an independent $U(0,1)$ random variable ;

(iii) Accept x^* if and only if $Z < f(x^*)$

$$\text{i.e.} \quad \text{iff} \qquad U \sup_x \frac{f(x)}{h(x)} < \frac{f(x^*)}{h(x^*)}. \tag{4.2.1}$$

The probability of rejection is

$$\frac{\int_{-\infty}^{\infty} [w(x) - f(x)]\, dx}{\int_{-\infty}^{\infty} w(x) dx} = 1 - \frac{1}{k} \tag{4.2.2}$$

Suppose that we assume a prior density for ξ, $\pi(\xi)$ and we take $\tilde{f}$ to be a posterior realisation of f. Now $\eta = \tilde{f}(\xi)$. Since $\tilde{f}$ is a monotonic function of ξ, the prior density of η is defined as

$$p(\eta) = \pi(\tilde{f}^{-1}(\eta)) \left| \frac{d\xi}{d\eta} \right| \tag{4.2.3}$$

The likelihood function associated with the m observations Y_j' is given by equation (4.1.3). The posterior density of η is given by

$$p(\eta|\mathbf{Y}') = cp(\eta)\mathcal{L}(\eta)$$

Suppose that M is the supremum of $\mathcal{L}(\eta)$ over the range of η, i.e. $\sup_\eta \mathcal{L}(\eta) = M$, then

$$\sup_\eta \frac{p(\eta|\mathbf{Y}')}{p(\eta)} = Mc = M^*. \tag{4.2.4}$$

This might suggest enveloping the posterior density of η by $M^* p(\eta)$ giving the following algorithm:

(i) Simulate ξ^* from the prior density $\pi(\xi)$;

(ii) Calculate $\eta^* = \tilde{f}(\xi^*)$;

(iii) Accept η^* (and hence ξ^*) iff $\mathcal{L}(\eta^*) \geq U \sup_\eta \mathcal{L}(\eta)$

where U is an independent $U(0,1)$ random variable. Using equation (4.2.2) the probability of rejection is given by $1 - \frac{1}{M^*}$. However this algorithm is likely to be most inefficient because typically M^* is large so the rejection probability will be high.

It was decided therefore to envelop the posterior distribution of η by $K^* p_u(\eta|\mathbf{Y}')$ where $p_u(\eta|\mathbf{Y}')$ is the posterior distribution of η which results from a uniform prior density for ξ. Thus K^* is defined as

$$\sup_\eta \frac{p(\eta|\mathbf{Y}')}{p_u(\eta|\mathbf{Y}')}$$

Theorem 4.1.

Suppose that ξ has prior density $\pi(\xi)$ and that this density is unimodal within $[x_1, x_n]$. Let

$$\pi_{\text{MAX}} = \max_{x_1 \le \xi \le x_n} \pi(\xi)$$

Realisations of the posterior distribution of $\eta, p(\eta|\mathbf{Y}')$ can be obtained by using the following algorithm:

(i) Simulate η^* from the posterior density $p_u(\eta|\mathbf{Y}')$;

(ii) Accept η^* as a realisation of a random variable with p.d.f. $p(\eta|\mathbf{Y}')$ iff

$$U\pi_{\text{MAX}} \le \pi\,(\tilde{f}^{-1}(\eta^*))$$

where $\tilde{f}$ is a posterior realisation of f and U is an independent $U(0,1)$ random variable.

Proof. The prior density $\pi(\xi)$ can be enveloped by $k\pi_u(\xi)$ where $\pi_u(\xi)$ denotes a uniform density on $[x_1, x_n]$.

$\pi_u(\xi)$ is defined as follows:

$$\pi_u(\xi) = \frac{1}{x_n - x_1} \qquad x_1 \le \xi \le x_n$$

Then $\exists\, k$ such that $\pi(\xi) \le k\pi_u(\xi)$, i.e.

$$k = \sup_{\xi} \frac{\pi(\xi)}{\pi_u(\xi)} = (x_n - x_1)\,\pi_{\text{MAX}} \tag{4.2.5}$$

since the supremum is attained for a ξ in the interval $[x_1, x_n]$.

Suppose the prior density for η corresponding to $\pi(\xi)$ is $p(\eta)$. The posterior density of η given $\mathbf{Y}'$ is given by

$$p(\eta|\mathbf{Y}') = C_p\, p(\eta)\, \mathcal{L}(\eta) \tag{4.2.6}$$

where C_p is a normalising constant and $\mathcal{L}(\eta)$ is given by (4.1.3). From equation (4.2.3)

$$p(\eta) = \pi(\tilde{f}^{-1}(\eta)) \left| \frac{d\xi}{d\eta} \right|$$

Let $p_u(\eta)$ be the prior density of η corresponding to the uniform density of ξ, $\pi_u(\xi)$

$$p_u(\eta) = \frac{1}{(x_n - x_1)} \left| \frac{d\xi}{d\eta} \right| \tag{4.2.7}$$

Therefore

$$\sup_{\tilde{f}(x_1)\le\eta\le\tilde{f}(x_n)} \frac{p(\eta)}{p_u(\eta)} = \sup_{x_1\le\xi\le x_n} \frac{\pi(\xi)}{\pi_u(\xi)} \cdot \frac{\left|\frac{d\xi}{d\eta}\right|}{\left|\frac{d\xi}{d\eta}\right|} = k$$

This means that

$$\frac{p(\eta)}{p_u(\eta)} \le k \qquad \eta\in[\tilde{f}(x_1),\tilde{f}(x_n)] \tag{4.2.8}$$

The posterior density of η, $p_u(\eta|\mathbf{Y}')$ is given by

$$p_u(\eta|\mathbf{Y}') = C_u p_u(\eta)\,\mathcal{L}(\eta). \tag{4.2.9}$$

Dividing (4.2.6) by (4.2.9) gives

$$\frac{p(\eta|\mathbf{Y}')}{p_u(\eta|\mathbf{Y}')} = \frac{C_p p(\eta)\mathcal{L}(\eta)}{C_u p_u(\eta)\mathcal{L}(\eta)} = \frac{C_p p(\eta)}{C_u p_u(\eta)} \tag{4.2.10}$$

Using equations (4.2.8) and (4.2.10)

$$\frac{p(\eta)}{p_u(\eta)} \le k \qquad \eta\in[\tilde{f}(x_1),\tilde{f}(x_n)]$$

$$\frac{p(\eta|\mathbf{Y}')}{p_u(\eta|\mathbf{Y}')} \le \frac{kC_p}{C_u} = k^* \qquad \eta\in[\tilde{f}(x_1),\tilde{f}(x_n)]$$

$$k^* = \sup_{\tilde{f}(x_1)\le\eta\le\tilde{f}(x_n)} \frac{p(\eta|\mathbf{Y}')}{p_u(\eta|\mathbf{Y}')}$$

It is possible therefore to envelop the posterior density function $p(\eta|\mathbf{Y}')$ by taking as the enveloping curve $k^* p_u(\eta|\mathbf{Y}')$. Using the rejection sampling algorithm (4.2.1) gives:

(i) Simulate η^* from $p_u(\eta|\mathbf{Y}')$;

(ii) Simulate $Z = Uk^* p_u(\eta^*|\mathbf{Y}')$ where U is a $U(0,1)$ random variable ;

(iii) Accept η^* iff $Z \le p(\eta^*|\mathbf{Y}')$.

Conditions (ii) and (iii) simplify since

$$Z = Uk^* p_u(\eta^*|\mathbf{Y}') = U(x_n - x_1)\pi_{\text{MAX}}\, C_p p_u(\eta^*)\mathcal{L}(\eta^*)$$

using equations (4.2.5) and (4.2.9). Using equation (4.2.6), condition (iii) becomes

$$\text{Accept } \eta^* \text{ iff } \quad U(x_n - x_1)\pi_{\text{MAX}}\, C_p p_u(\eta^*)\mathcal{L}(\eta^*) \le C_p p(\eta^*)\mathcal{L}(\eta^*)$$

This simplifies to

$$\text{Accept } \eta^* \text{ iff } \quad U(x_n - x_1)\pi_{\text{MAX}}\, p_u(\eta^*) \le p(\eta^*).$$

Now dividing equation (4.2.3) by equation (4.2.7) gives

$$\frac{p(\eta^*)}{p_u(\eta^*)} = \pi(\tilde{f}^{-1}(\eta^*))\,(x_n - x_1)$$

So condition (iii) becomes

$$\text{Accept } \eta^* \text{ iff}$$

$$U\pi_{MAX} \leq \pi(\tilde{f}^{-1}(\eta^*))$$

4.3 INTRODUCTION TO A SWITCHING ALGORITHM

The above theorem has shown that it is relatively easy to simulate from the posterior density $p(\eta|\mathbf{Y}')$ which corresponds to a prior density for $\xi, \pi(\xi)$, provided one can obtain a posterior realisation from the density $p_u(\eta|\mathbf{Y}')$, the posterior density of η corresponding to a uniform prior density for ξ. The question now arises as to how to efficiently simulate from the posterior density $p_u(\eta|\mathbf{Y}')$.

The method which is put forward in this section is an adaptation of a switching algorithm derived by Atkinson and Whittaker (1976). They suggested a switching algorithm for the efficient generation of Beta random variables with at least one parameter less than 1. The algorithm is used for densities of the product form

$$g(x) = kg_1(x)g_2(x)$$

where k is a constant. The posterior density $p_u(\eta|\mathbf{Y}')$ has this form since it is defined as

$$p_u(\eta|\mathbf{Y}') = C_u p_u(\eta)\mathcal{L}(\eta) \tag{4.3.1}$$

where $\tilde{f}(x_1) \leq \eta \leq \tilde{f}(x_n)$. The algorithm was derived for $0 \leq x \leq 1$ so to facilitate the modification of the algorithm to simulate from $p_u(\eta|\mathbf{Y}')$, a linear transformation of the form $\zeta = a\eta + b$ was made first so that $0 \leq \zeta \leq 1$. If $\tilde{f}$ is monotonic increasing then a and b are defined as follows:

$$a = \frac{1}{\tilde{f}(x_n)-\tilde{f}(x_1)} \quad b = \frac{-\tilde{f}(x_1)}{\tilde{f}(x_n)-\tilde{f}(x_1)} .$$

If $\tilde{f}$ is monotonic decreasing then a and b are defined as

$$a = \frac{1}{\tilde{f}(x_1)-\tilde{f}(x_n)} \quad b = \frac{-\tilde{f}(x_n)}{\tilde{f}(x_1)-\tilde{f}(x_n)} .$$

Using (4.1.3) the posterior density $\pi_u(\zeta|\mathbf{Y}')$ is given by

$$\pi_u(\zeta|\mathbf{Y}') = k_u \pi_u(\zeta)\mathcal{L}(\zeta)$$

where $\mathcal{L}(\zeta)$ is the likelihood function given by

$$\mathcal{L}(\zeta) = \left[\frac{m}{2\pi a^2\sigma^2}\right]^{\frac{1}{2}} \exp\left[-\frac{m}{2a^2\sigma^2}(\zeta - a\bar{Y}' - b)^2\right] \tag{4.3.2}$$

Here $\bar{Y}'$ is the mean of the m observations at the prediction stage of the calibration process.

$$\pi_u(\zeta) = p_u\left[\frac{\zeta-b}{a}\right]\left|\frac{d\eta}{d\zeta}\right| = \frac{1}{a}\,p_u\left[\frac{\zeta-b}{a}\right] \tag{4.3.3}$$

Let $\pi_u(\zeta) = g_1(\zeta)$ and $\mathcal{L}(\zeta) = g_2(\zeta)$ so that the posterior density $\pi_u(\zeta|\mathbf{Y}')$ is given by

$$\pi_u(\zeta|\mathbf{Y}') = k_u\, g_1(\zeta)\, g_2(\zeta)$$

Let U be a uniform random variable on $[0, 1]$. If one wished to simulate from the posterior density $\pi_u(\zeta|\mathbf{Y}')$ using rejection sampling *only*, one could either

(i) simulate ζ^* from $g_1(\zeta)$

(ii) accept ζ^* iff

$$\frac{g_2(\zeta^*)}{\sup\limits_{0\le\zeta\le1} g_2(\zeta)} \ge U$$

or one could

(iii) simulate ζ^* from $g_2(\zeta)$

(iv) accept ζ^* iff

$$\frac{g_1(\zeta^*)}{\sup\limits_{0\le\zeta\le1} g_1(\zeta)} \ge U$$

Provided both supremums were finite, ζ^* would have density $\pi_u(\zeta|\mathbf{Y}')$. The algorithm of Atkinson and Whittaker however is a mixture of two sampling methods; rejection and composition.

The idea of the algorithm is to break the range of ζ into two parts $[0,t]$ and $[t,1]$ and to sample from $g_1(\zeta)$ with probability θ if $\zeta \in [0,t]$ and to sample from $g_2(\zeta)$ with probability $(1-\theta)$ if $\zeta \in [t,1]$. As both g_1 and g_2 are bounded there is a choice, either we can take

$$\text{(a)}\quad g_1(\zeta) = \pi_u(\zeta) \quad g_2(\zeta) = \mathcal{L}(\zeta)$$

or we can take

$$\text{(b)}\quad g_1(\zeta) = \mathcal{L}(\zeta) \quad g_2(\zeta) = \pi_u(\zeta).$$

The choice of which option (a) or (b) and the choice of t were based on optimising the efficiency of the algorithm for the case of $\tilde{f} = \hat{f}_\alpha$ where $\hat{f}_\alpha$ is the spline smoother of the calibration data and this will be discussed in more detail in

Section 4.5 . It is only necessary to say at this stage that t was chosen to correspond after transformation to one of the knots x^* where $x^* \in \{x_i\}$ $i=2,3,\ldots,(n-1)$. So

$$t = a\tilde{f}(x^*) + b \ . \tag{4.3.4}$$

Suppose that for the rest of the discussion it is assumed that $\tilde{f}$ is monotonic increasing where $\tilde{f}$ is a posterior realisation of f.

4.4 DETAILS OF THE SWITCHING ALGORITHM

Suppose that option (a) is chosen and let

$$G_1(x) = \int_0^x g_1(y)dy \quad \text{and} \quad G_2(x) = \int_0^x g_2(y)dy.$$

Let U_1 and U_2 be independent uniform random variables on $[0,1]$ and break the range of ζ into two parts $[0,t]$ and $[t,1]$ where $0 < t < 1$. With probability θ

$$G_1(\zeta) = U_1 G_1(t)$$

$$\zeta = G_1^{-1}(U_1 G_1(t)) \tag{4.4.1}$$

and ζ is accepted iff

$$\frac{g_2(\zeta)}{\sup_{0\le\zeta\le t} g_2(\zeta)} \ge U_2 \tag{4.4.2}$$

while with probability $(1-\theta)$

$$G_2(\zeta) = [G_2(1)-(1-U_1)\ \{G_2(1)-G_2(t)\}]$$

$$\zeta = G_2^{-1}[G_2(1)-(1-U_1)\ \{G_2(1)-G_2(t)\}] \tag{4.4.3}$$

This is accepted iff

$$\frac{g_1(\zeta)}{\sup_{t\le\zeta\le 1} g_1(\zeta)} \ge U_2 \tag{4.4.4}$$

The probability θ depends on t and is given by

$$\frac{\theta}{(1-\theta)} = \frac{G_1(t)}{\sup_{t\le\zeta\le 1} g_1(\zeta)} \Big/ \frac{G_2(1)-G_2(t)}{\sup_{0\le\zeta\le t} g_2(\zeta)} \tag{4.4.5}$$

Equations (4.4.1), (4.4.3) and (4.4.5) and inequalities (4.4.2) and (4.4.4) define the switching algorithm for generating values from the posterior density

$$\pi_u(\zeta|\mathbf{Y}') = k_u\, g_1(\zeta) g_2(\zeta) \ .$$

4.4.1 Consideration of G_1

$G_1(\zeta)$ is the distribution function associated with the prior density $\pi_u(\zeta)$ since

$$G_1(x) = \int_0^x g_1(y)dy = \int_0^x \pi_u(y)dy$$

In order to sample from the prior density $\pi_u(\zeta)$ over $[0,t]$ we require that

$$G_1(\zeta) = U_1 G_1(t)$$

using equation (4.4.1) Assume that ξ has a uniform prior density on $[x_1, x_n]$ so that

$$\pi(\xi) = \frac{1}{(x_n - x_1)} \qquad x_1 \le \xi \le x_n .$$

The distribution function of ξ associated with this density is given by

$$P(\xi \le x^*) = \frac{x^* - x_1}{(x_n - x_1)} .$$

As $\eta = \tilde{f}(\xi)$ is a monotonic increasing function of ξ then the distribution function of η is given by

$$P(\eta \le \eta^*) = \frac{\tilde{f}^{-1}(\eta^*) - x_1}{(x_n - x_1)} \quad \text{where } \eta^* = \tilde{f}(x^*) .$$

From equation (4.3.4) $t = a\tilde{f}(x^*) + b$. The distribution function of ζ, G_1 is therefore given by

$$G_1(t) = P(\zeta \le t) = \frac{\tilde{f}^{-1}(\frac{t-b}{a}) - x_1}{(x_n - x_1)}$$

So, in order to simulate ζ from the prior density $\pi_u(\zeta)$ over the range $0 \le \zeta \le t$ we require

$$\frac{\tilde{f}^{-1}(\frac{\zeta-b}{a}) - x_1}{(x_n - x_1)} = U_1 \frac{(x^* - x_1)}{(x_n - x_1)}$$

$$\text{i.e. } \zeta = a\tilde{f}(x_1 + U_1(x^* - x_1)) + b . \qquad (4.4.6)$$

4.4.2 Consideration of the supremum of $g_2(\zeta)$ for $0 \leq \zeta \leq t$

The value ζ simulated from the prior density $\pi_u(\zeta)$ using equation (4.4.6) is acceptable iff

$$\frac{g_2(\zeta)}{\sup_{0\leq\zeta\leq t} g_2(\zeta)} \geq U_2$$

from inequality (4.4.2), so consider the calculation of

$$\sup_{0\leq\zeta\leq t} g_2(\zeta) = \sup_{0\leq\zeta\leq t} \mathcal{L}(\zeta)$$

From equation (4.3.2) $\mathcal{L}(\zeta)$ is given by

$$\left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}} \exp\left[\frac{-m}{2a^2\sigma^2}(\zeta - a\bar{Y}' - b)^2\right].$$

$$\sup_{0\leq\zeta\leq t} g_2(\zeta) = \left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}} \sup_{0\leq\zeta\leq t} \exp\left[\frac{-m}{2a^2\sigma^2}(\zeta - a\bar{Y}' - b)^2\right]$$

$$= \frac{1}{a} \sup_{\tilde{f}(x_1)\leq\eta\leq\tilde{f}(x^*)} \psi(\eta)$$

where

$$\psi(\eta) = \left(\frac{m}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[\frac{-m}{2\sigma^2}(\eta - \bar{Y}')^2\right].$$

The shape of $\psi(\eta)$ is that of a normal curve symmetrical about $\bar{Y}'$. If $\tilde{f}(x_1) \leq \bar{Y}' \leq \tilde{f}(x^*)$ then the supremum occurs at $\eta = \bar{Y}'$ so

$$\sup_{0\leq\zeta\leq t} g_2(\zeta) = \left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}}$$

$$\text{Let } z_n = \frac{\sqrt{m}(\tilde{f}(x_n) - \bar{Y}')}{\sigma} \quad \text{and} \quad z_1 = \frac{\sqrt{m}(\tilde{f}(x_1) - \bar{Y}')}{\sigma} \tag{4.4.7}$$

If $\bar{Y}' < \tilde{f}(x_1)$ then the supremum of $\psi(\eta)$ occurs at $\eta = \tilde{f}(x_1)$ so

$$\sup_{0\leq\zeta\leq t} g_2(\zeta) = \left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}z_1^2\}$$

whereas if $\bar{Y}' > \tilde{f}(x^*)$ then the supremum of $\psi(\eta)$ occurs at $\eta = \tilde{f}(x^*)$ so

$$\sup_{0\leq\zeta\leq t} g_2(\zeta) = \left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}z^{*2}\}$$

where $z^* = \dfrac{\sqrt{m}(\tilde{f}(x^*) - \bar{Y}')}{\sigma}$.

4.4.3 Consideration of G_2

The function $G_2(\zeta)$ is defined as

$$\int_0^{\zeta} \mathcal{L}(y)\,dy = \left(\frac{m}{2\pi a^2\sigma^2}\right)^{\frac{1}{2}} \int_0^{\zeta} \exp\left[\frac{-m}{2a^2\sigma^2}\,(y-a\bar{Y}'-b)^2\right] dy$$

$$= \Phi(z_\zeta) - \Phi(z_1)$$

where $z_\zeta = \dfrac{\sqrt{m}\,(\zeta-a\bar{Y}'-b)}{a\sigma}$, z_1 is as defined in equation (4.4.7) and Φ is the standardised normal distribution function. In order to sample from $\mathcal{L}(\zeta)$ over $[t,1]$ we require that

$$G_2(\zeta) = [G_2(1)-(1-U_1)\{G_2(1)-G_2(t)\}] \tag{4.4.8}$$

Equation (4.4.8) thus becomes

$$\Phi(z_\zeta) - \Phi(z_1) = [\Phi(z_n) - \Phi(z_1) - (1-U_1)[\Phi(z_n) - \Phi(z_1) - \{\Phi(z_t) - \Phi(z_1)\}]$$

$$\text{i.e.}\quad \Phi(z_\zeta) = \Phi(z_n) - (1-U_1)\{\Phi(z_n) - \Phi(z_t)\}$$

$$\text{where}\quad z_t = \frac{\sqrt{m}\,(t-a\bar{Y}'-b)}{a\sigma} = \frac{\sqrt{m}\,(\tilde{f}(x^*)-\bar{Y}')}{\sigma}. \tag{4.4.9}$$

It follows that

$$z_\zeta = \Phi^{-1}[\Phi(z_n)-(1-U_1)\{\Phi(z_n)-\Phi(z_t)\}]$$

$$\zeta = z_\zeta\,\frac{a\sigma}{\sqrt{m}} + a\bar{Y}' + b$$

$$= \frac{a\sigma}{\sqrt{m}}\Phi^{-1}[\Phi(z_n)-(1-U_1)\{\Phi(z_n)-\Phi(z_t)\}] + a\bar{Y}' + b \tag{4.4.10}$$

4.4.4 Consideration of the supremum of $g_1(\zeta)$ for $t \le \zeta \le 1$

The value of ζ simulated from $\mathcal{L}(\zeta)$ using equation (4.4.10) is acceptable iff

$$\frac{g_1(\zeta)}{\sup_{t\le\zeta\le1} g_1} \ge U_2$$

from inequality (4.4.4). Let us now consider the calculation of $\sup_{t\le\zeta\le1} g_1 = \sup_{t\le\zeta\le1} \pi_u(\zeta)$.

Using equations (4.2.3), (4.3.3) and (4.3.4) we obtain the following

$$\sup_{t\le\zeta\le1} \pi_u(\zeta) = \frac{1}{a} \sup_{x^*\le\xi\le x_n} \pi(\xi)\,\frac{d\xi}{d\eta}$$

where $\pi(\xi)$ is the prior density of ξ. This prior density is assumed to be uniform

so the above equation becomes

$$\sup_{t\le\zeta\le 1} \pi_u(\zeta) = \frac{1}{a}\,\frac{1}{(x_n - x_1)}\,\sup_{x^*\le\xi\le x_n}\left(\frac{1}{\frac{d\eta}{d\xi}}\right)$$

Now $\eta = \tilde{f}(\xi)$ where $\tilde{f}$ is a posterior realisation of f. Now for $x_i \le \xi < x_{i+1}$ $i=1,2,\ldots n-1$

$$\eta = a_i(\xi - x_i)^3 + b_i(\xi - x_i)^2 + c_i(\xi - x_i) + d_i$$

$$\text{so} \quad \frac{d\eta}{d\xi} = 3a_i(\xi - x_i)^2 + 2b_i(\xi - x_i) + c_i = \Delta_\xi$$

which is a quadratic in ξ. In evaluating the supremum over $[x^*, x_n]$ of $\frac{1}{\Delta_\xi}$ there are various cases to consider;

Case (i). The quadratic function Δ_ξ has a maximum which is attained in the interval $[x_i, x_{i+1})$ i.e. $a_i < 0$ so that the supremum is given by

$$\sup_{x^*\le\xi\le x_n} \frac{1}{\Delta_\xi} = \frac{1}{\min(c_i, c_{i+1})} \qquad i=i^*,(i^*+1),(i^*+2)\ldots.(n-1)$$

where $x_{i^*} = x^*$.

Case (ii). The quadratic function Δ_ξ has a minimum which is attained within the interval $[x_i, x_{i+1})$ i.e. $a_i > 0$ and $b_i^2 \le 3a_ic_i$. Here the supremum is given by

$$\sup_{x^*\le\xi\le x_n} \frac{1}{\Delta_\xi} = \frac{3a_i}{(3a_ic_i - b_i^2)} \qquad i=i^*,(i^*+1),(i^*+2)\ldots.(n-1)$$

Case (iii). The quadratic function Δ_ξ has a minimum which is attained in an interval below $[x_i, x_{i+1})$ i.e. $a_i > 0$ and $b_i > 0$. Here the supremum is given by

$$\sup_{x^*\le\xi\le x_n} \frac{1}{\Delta_\xi} = \frac{1}{c_i} \qquad i=i^*,(i^*+1),(i^*+2)\ldots.(n-1)$$

Case (iv). The quadratic function Δ_ξ has a minimum which is achieved in an interval above $[x_i, x_{i+1})$, i.e. $a_i > 0$ and $x_i - \frac{b_i}{3a_i} > x_{i+1}$. Here the supremum is given by

$$\sup_{x^*\le\xi\le x_n} \frac{1}{\Delta_\xi} = \frac{1}{c_{i+1}} \qquad i=i^*,(i^*+1),(i^*+2)\ldots.(n-1)$$

4.4.5 Final form of algorithm

Rather than generate an independent uniform to determine the switch, i.e. whether one samples from $g_1(\zeta)$ or whether one samples from $g_2(\zeta)$, Atkinson and Whittaker (1976) suggested using U_1 in the following way. If $U_1 < \theta$ then replace U_1 in equation (4.4.1) by U_1/θ and if $U_1 > \theta$ then replace $(1-U_1)$ in equation (4.4.3) by $(1-U_1)/(1-\theta)$. If one takes logarithms to the base e, inequality (4.4.2) becomes

$$\ln\{\sup_{0\le\zeta\le t} g_2(\zeta)\} - \ln g_2(\zeta) \le E$$

where E is an exponential random variable with mean 1. This follows because $X = -\frac{1}{\lambda}\ln U$, where U is an independent $U(0,1)$ random variable, has an exponential distribution with parameter λ.

Similarly, inequality (4.4.4) becomes

$$\ln\{\sup_{t\le\zeta\le 1} g_1(\zeta)\} - \ln g_1(\zeta) \le E.$$

The algorithm using option (a) becomes:

(1) Choose x^* and calculate t and θ using equations (4.3.4) and (4.4.5).

(2) Generate U and E. If $U > \theta$ go to (4).

(3) Set $\zeta = a\tilde{f}(x_1+\frac{U}{\theta}(x^*-x_1)) + b$.

If $\ln\{\sup_{0\le\zeta\le t} L(\zeta)\} - \ln L(\zeta) \le E$ accept ζ, otherwise go to (2).

(4) Set $\zeta = \frac{a\sigma}{\sqrt{m}}\Phi^{-1}[\Phi(z_n) - (\frac{1-U}{1-\theta})\{\Phi(z_n)-\Phi(z_t)\}] + a\bar{Y}' + b$

where z_1, z_n and z_t are given by equations (4.4.7) and (4.4.9).

If $\ln\{\sup_{t\le\zeta\le 1} \pi_u(\zeta)\} - \ln\pi_u(\zeta) \le E$ accept ζ, otherwise go to (2).

When option (b) is chosen, the steps of the algorithm are as follows:

(1) As step (1) above.

(2) As step (2) above.

(3) Set $\zeta = \frac{a\sigma}{\sqrt{m}}\Phi^{-1}[(1-\frac{U}{\theta})\Phi(z_1) + \frac{U}{\theta}\Phi(z_t)] + a\bar{Y}' + b$

where z_1 and z_t are given by equations (4.4.7) and (4.4.9)

If $\ln\{\sup_{0\le\zeta\le t} \pi_u(\zeta)\} - \ln\pi_u(\zeta) \le E$ accept ζ, otherwise go to (2).

(4) Set $\zeta = a\tilde{f}(x_n - \frac{(1-U)}{(1-\theta)}(x_n - x^*)) + b$.

If $\ln\{\sup_{t\le\zeta\le 1} L(\zeta)\} - \ln L(\zeta) \le E$ accept ζ, otherwise go to (2).

Algorithms similar to those above were also derived for the case of $\tilde{f}$ being monotonic decreasing.

4.5 CONSIDERATIONS OF EFFICIENCY

The efficiency of the switching algorithm is defined as being the probability of accepting a value of ζ. Suppose the posterior density $\pi_u(\zeta|\mathbf{Y}')$ was enveloped by $kg_1(\zeta)$ over the whole interval $[0,1]$ where k is given by

$$k = \sup_{\zeta} \frac{\pi_u(\zeta|\mathbf{Y}')}{g_1(\zeta)} = k_u \sup_{\zeta} g_2(\zeta)$$

then the probability of accepting a value of ζ would be $1/\{k_u \sup_{\zeta} g_2(\zeta)\}$. The switching algorithm is a mixture of composition and rejection sampling so the probability of accepting a value of ζ is

$$\frac{\theta}{\int_0^t \{k_u \sup_{0\le\zeta\le t} g_2(\zeta)\}\, g_1(\zeta)d\zeta} = \frac{\theta}{\{k_u \sup_{0\le\zeta\le t} g_2(\zeta)\}\, G_1(t)}$$

or equivalently

$$\frac{(1-\theta)}{\int_t^1 \{k_u \sup_{t\le\zeta\le 1} g_1(\zeta)\}\, g_2(\zeta)d\zeta} = \frac{(1-\theta)}{\{k_u \sup_{t\le\zeta\le 1} g_1(\zeta)\}[G_2(1)-G_2(t)]}$$

since we sample from $g_1(\zeta)$ with probability θ if $0 \le \zeta \le t$ and sample from $g_2(\zeta)$ with probability $(1-\theta)$ if $t < \zeta \le 1$. Use of equation (4.4.5) will verify that the two probabilities of acceptance given above are equal. Using equation (4.4.5) the probability θ is given by

$$\theta = \frac{G_1(t) \sup_{0\le\zeta\le t} g_2(\zeta)}{[G_2(1)-G_2(t)] \sup_{t\le\zeta\le 1} g_1(\zeta) + G_1(t) \sup_{0\le\zeta\le t} g_2(\zeta)} \quad . \qquad (4.5.1)$$

The equation for the efficiency of the algorithm is therefore given by

$$\text{Efficiency} = \frac{1}{k_u\{[G_2(1)-G_2(t)] \sup_{t\le\zeta\le 1} g_1(\zeta) + G_1(t) \sup_{0\le\zeta\le t} g_2(\zeta)\}}$$

Atkinson and Whittaker (1976) were able to obtain an analytically closed form for the efficiency of the procedure and they then differentiated this with respect to t to obtain the value of t, t_{opt}, which maximised the efficiency of the procedure. This is not possible here because of the form of G_1 and the supremums (see

Sections 4.4.1, 4.4.2 and 4.4.4). It was considered difficult and time-consuming to try and optimise the efficiency for *every* posterior realisation $\tilde{f}$. It was therefore decided to consider only the efficiency for $\tilde{f} = \hat{f}_\alpha$ where $\hat{f}_\alpha$ is the spline smoother for the calibration data set. It may be recalled that $\hat{f}_\alpha$ is the mode of the posterior distribution of f. The efficiency was evaluated for the $(n-2)$ values of t given by

$$t_i = a\hat{f}_\alpha(x_i) + b \qquad i=2,3,\ldots(n-1)$$

and for both options (a) and (b). With option (a),

$$g_1(\zeta) = \pi_u(\zeta) \text{ and } g_2(\zeta) = \mathcal{L}(\zeta)$$

whereas with option (b)

$$g_1(\zeta) = \mathcal{L}(\zeta) \text{ and } g_2(\zeta) = \pi_u(\zeta).$$

The value of θ was also evaluated for each different value of t. Suppose the knot corresponding to the highest value of efficiency is x^*. The knot x^* was chosen to be that knot which maximised the efficiency subject to $0.001 \leq \theta \leq 0.999$. This constraint on θ was necessary to ensure the possibility of sampling from *both* g_1 and g_2. Since the posterior density $\pi_u(\zeta|\mathbf{Y}')$ varies with $\mathbf{Y}'$, it was necessary to determine x^* for each different value of $\bar{Y}'$. The option chosen, (a) or (b), was that option which gave the highest value for the efficiency, for a given x^* and $\bar{Y}'$.

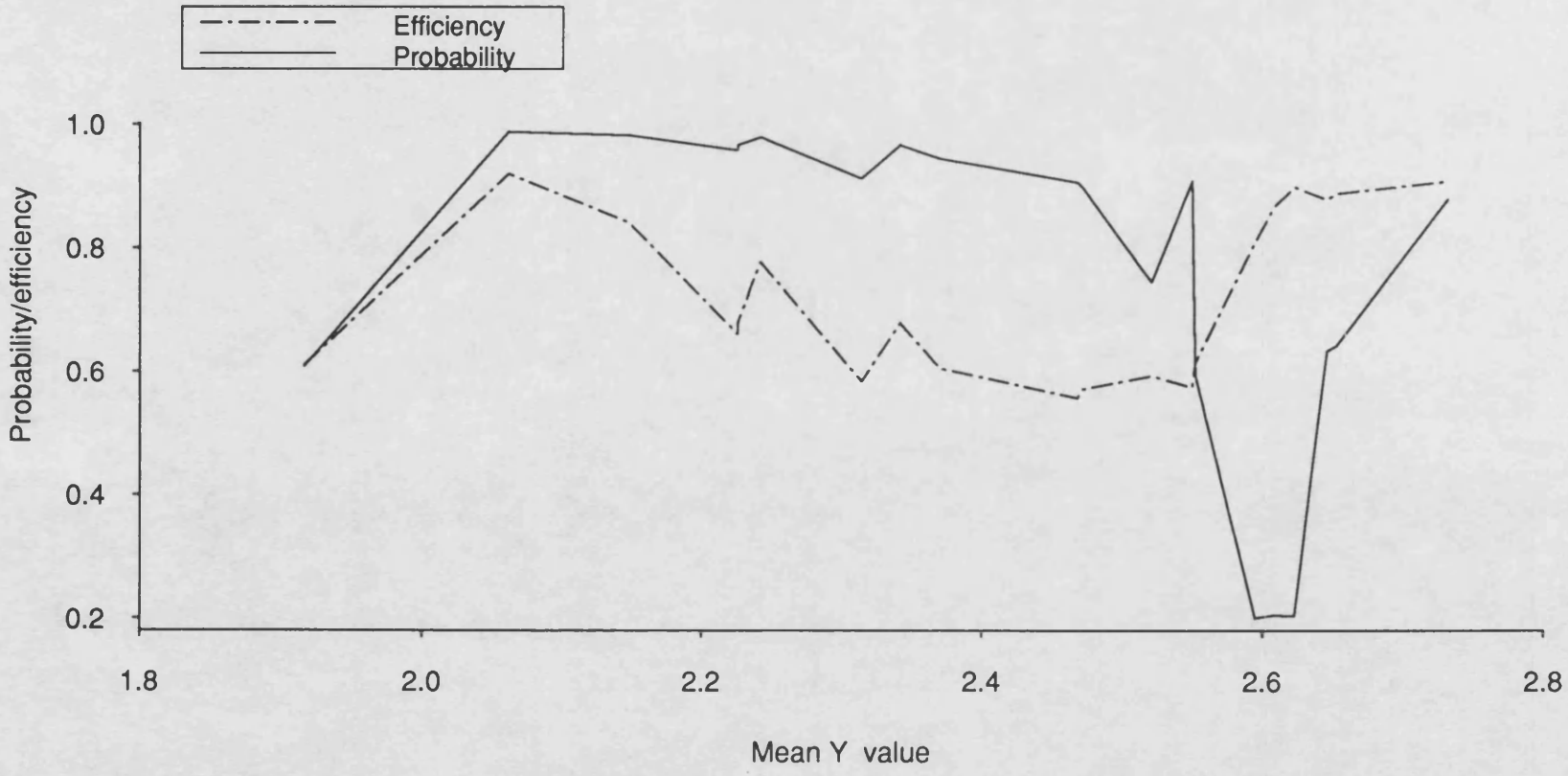
The best option for all the simulated data sets, except the Gompertz data set, was option (b). Fig. 4.1 shows θ plotted against $\bar{Y}'$ for the Asymptotic data set where θ = Probability of sampling from $\mathcal{L}(\zeta)$ over $[0,t]$ = Probability of sampling from $\mathcal{L}(\eta)$ over $[\hat{f}(x_1), \hat{f}(x^*)]$. For this data set, the constraint $0.001 \leq \theta \leq 0.999$ was not binding, producing a unique *optimal knot* x^* for a given $\bar{Y}'$. The corresponding efficiencies are also given in the same figure. The best option for the Gompertz data set was option (a). With some data sets, e.g. antibiotic assay data, upper central and upper canine teeth data, both options were used to simulate posterior realisations of ζ and η.

It is interesting to note, by way of comparison, that the sampling method used in Chapter 3 (Section 3.2.2) to simulate values of η from the posterior density $p(\eta|\mathbf{Y}')$ is equivalent to sampling from $\mathcal{L}(\zeta)$ over the whole interval $[0,1]$ with probability 1. This is because it was assumed in Section 3.2.2 that η had a uniform prior density on $[\tilde{f}(x_1), \tilde{f}(x_n)]$ where $\tilde{f}$ is a posterior realisation of f. Hence the posterior density $p(\eta|\mathbf{Y}')$ is given by

$$p(\eta|\mathbf{Y}') \propto \mathcal{L}(\eta)$$

where $\mathcal{L}(\eta)$ is given by equation (4.1.3). Therefore, simulating from the posterior

Fig.4.1: Probabilities and efficiencies (Asymptotic data set)



density $p(\eta|\mathbf{Y}')$ is equivalent to simulating from $\mathcal{L}(\eta)$. Since $\zeta = a\eta + b$ where a and b are defined in Section 4.3

$$\pi(\zeta|\mathbf{Y}') \propto \mathcal{L}(\zeta)$$

where $\mathcal{L}(\zeta)$ is given by equation (4.3.2). So to simulate posterior realisations of ζ one simulates from $\mathcal{L}(\zeta)$ over $[0,1]$ with probability 1.

4.6 APPLICATION TO THE SIMULATED DATA SETS

The methods derived in Section 4.2 and Section 4.4 enable one to obtain realisations of $\pi(\xi|\mathbf{Y}')$ for any given prior density for ξ. It was decided to test out the methods on these data sets assuming three different priors for ξ, namely

(i) a uniform density on $[x_1, x_n]$

(ii) a triangular density on $[x_1, x_n]$

(iii) a truncated non-central Student density

The six calibration data sets analysed here are the same six data sets as those considered in Section 3.3.1, Chapter 3. For each of these data sets the total number of observations, N, is 240. To assess the performance of this method three observations were simulated at each of twenty newly chosen X-values covering the calibration range. So the model for the prediction stage of the calibration process is given by

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3.$$

Posterior distributions $\pi(\xi|\mathbf{Y}')$ were constructed using the methods described below. The posterior median was again used to give a point estimate for ξ and interval estimates for ξ were obtained in the same way as described in Chapter 3, Section 3.3.1. Interval estimates were termed *successful* if the true value ξ lay within the constructed interval.

4.6.1 A uniform prior density

This is the easiest prior to work with because only the switching algorithms laid out in Section 4.4.5 are needed. With the other priors it is necessary to use in addition the results of Theorem 4.1.

(1) A posterior realisation of f, $\tilde{f}_s$, was obtained using the method described in Section 3.2.2, Chapter 3.

(2) Equation (4.3.4) was used to calculate t for a given x^* and $\tilde{f}_s$.

(3) The value of θ was calculated using equation (4.5.1). If this probability was greater than 0.999 or less than 0.001, another posterior realisation of f was generated and steps (2) – (3) repeated.

(4) Using this value of θ and t, realisations of $\pi_u(\zeta|\mathbf{Y}')$ were obtained where $\pi_u(\zeta|\mathbf{Y}')$ is the posterior density corresponding to a uniform prior density for ξ. This was achieved by using the algorithms laid out in Section 4.4.5. Suppose a posterior realisation of ζ is denoted by ζ_v^*.

(5) The values of

$$\tilde{f}_s^{-1}\left(\frac{\zeta_v^*-b}{a}\right) \qquad \begin{array}{l} s=1,2,3,\ldots,199,\ v=1,2,\ldots,50 \\ s=200,\ v=1,2,3\ldots.49 \end{array}$$

were calculated giving 9999 posterior realisations of ξ. It should be noted that a *different* set of 50 realisations of $\pi_u(\zeta|\mathbf{Y}')$ were generated for each spline, $\tilde{f}_s$.

Table 4.1 shows the calibration range and mean absolute error for each of the six data sets. Table 4.2 shows the maximum width, minimum width, and mean width of the twenty interval estimates together with the proportion of successful estimates. The probability level for the interval estimates is 90%.

4.6.2 A triangular prior density

Let $\frac{1}{2}(x_1+x_n) = x_m$. The triangular density function is defined as follows:

$$\begin{aligned} \pi(\xi) &= \frac{4(\xi-x_1)}{(x_n-x_1)^2} && x_1 \le \xi \le x_m \\ &= \frac{4(x_n-\xi)}{(x_n-x_1)^2} && x_m \le \xi \le x_n \\ &= 0 && \text{otherwise.} \end{aligned} \qquad (4.6.1)$$

To generate realisations of $\pi(\xi|\mathbf{Y}')$, where the prior density for ξ is a triangular density, it is first necessary to generate realisations of $\pi_u(\zeta|\mathbf{Y}')$. This is achieved by performing steps (1) – (4) of Section 4.6.1. The next stage is to obtain a realisation of $\pi_{Tr}(\zeta|\mathbf{Y}')$ which is the posterior density of ζ corresponding to the triangular prior density $\pi(\xi)$. This is achieved by applying Theorem 4.1. Suppose such a realisation of $\pi_{Tr}(\zeta|\mathbf{Y}')$ is called ζ_{Tr}. The last stage of the procedure is to obtain a realisation of $\pi_{Tr}(\xi|\mathbf{Y}')$, ξ_{Tr}. This is achieved by taking

$$\xi_{Tr} = \tilde{f}_s^{-1}\left(\frac{\zeta_{Tr}-b}{a}\right)$$

where a and b are defined in Section 4.3. and $\tilde{f}_s$ is a posterior realisation of f.

Table 4.1 Point estimates for ξ assuming a uniform prior for ξ

Data set	Calibration range	Mean absolute error
Gompertz	0.00 - 13.43	0.35
Weibull	10.00 - 79.52	0.71
Preece-Baines	4.00 - 19.80	0.15
Bleasdale-Nelder	20.00 - 178.00	3.18
Asymptotic	1.00 - 21.74	1.10
Linear	15.00 - 70.00	3.59

Table 4.2 Interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Data set	Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Gompertz	2.53	0.67	1.57	19/20
Weibull	8.09	1.93	3.55	19/20
Preece-Baines	2.69	0.17	0.74	19/20
Bleasdale-Nelder	26.64	4.06	13.25	18/20
Asymptotic	8.40	1.16	4.95	19/20
Linear	16.84	10.38	15.27	19/20

According to Theorem 4.1 given in Section 4.2, a realisation $p(\eta|\mathbf{Y}')$ corresponding to a prior density for ξ, $\pi(\xi)$, is obtained by simulating η^* from $p_u(\eta|\mathbf{Y}')$ using the switching algorithms and accepting η^* if and only if

$$U\,\pi_{\text{MAX}} \leq \pi(\tilde{f}^{-1}(\eta^*))$$

where $\pi_{\text{MAX}} = \underset{x_1 \leq \xi \leq x_n}{\text{Max}}\ \pi(\xi)$ and $\tilde{f}$ is a posterior realisation of f. For the triangular density defined in equation (4.6.1) the maximum occurs at $\xi = x_m$ so π_{MAX} is given by

$$\pi_{\text{MAX}} = \frac{2}{(x_n - x_1)}$$

Now $\eta = a\zeta + b$ so using ζ rather than η, one obtains a realisation of $\pi_{Tr}(\zeta|\mathbf{Y}')$ by

(a) Simulating ζ^* from $\pi_u(\zeta|\mathbf{Y}')$

(b) Accepting ζ^* iff

$$U\,\pi_{\text{MAX}} \leq \pi(\tilde{f}^{-1}(\frac{\zeta^* - b}{a}))$$

i.e. if and only if

$$U \leq 2\frac{(\xi^* - x_1)}{(x_n - x_1)} \qquad \text{if } x_1 \leq \xi^* \leq x_m$$

$$U \leq 2\frac{(x_n - \xi^*)}{(x_n - x_1)} \qquad \text{if } x_m \leq \xi^* \leq x_n \qquad (4.6.2)$$

where $\xi^* = \tilde{f}^{-1}\,(\frac{\zeta^* - b}{a})$ and U is an independent $U(0,1)$ random variable.

Using the algorithm (4.6.2), 50 realisations of $\pi_{Tr}(\zeta|\mathbf{Y}')$ were obtained. Suppose these are denoted by ζ_v^* (v=1,2,3.....50). For each posterior realisation of f, $\tilde{f}_s$, the values of

$$\tilde{f}_s^{-1}\,(\frac{\zeta_v^* - b}{a}) \qquad \begin{array}{l} s=1,2,3,\ldots,199,\ v=1,2,\ldots,50 \\ s=200,\ v=1,2,3\ldots.49 \end{array}$$

were calculated giving 9999 posterior realisations of ξ. By way of illustration, Table 4.3 gives the calibration range and mean absolute errors for the Gompertz, Preece-Baines and Bleasdale-Nelder data sets. Table 4.4 gives the maximum width, the minimum width and mean width of the twenty interval estimates for ξ together with the proportion of successful estimates. The probability level for the interval estimates is 90%. Let $\hat{\xi}_{Tr}$ be the point estimate of ξ when one assumes a triangular prior density for ξ and let $\hat{\xi}_u$ be the point estimate of ξ when one

Table 4.3 Point estimates for ξ assuming a triangular prior for ξ

Data set	Calibration range	Mean absolute error
Gompertz	0.00 - 13.43	0.40
Preece-Baines	4.00 - 19.80	0.13
Bleasdale-Nelder	20.00 - 178.00	3.12

Table 4.4 Interval estimates for ξ assuming a triangular prior for ξ (90% probability level)

Data set	Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Gompertz	2.07	0.68	1.44	18/20
Preece-Baines	2.38	0.20	0.70	18/20
Bleasdale-Nelder	23.84	3.84	12.70	17/20

assumes a uniform prior density for ξ. Fig. 4.2 shows $\hat{\xi}_{Tr}$ and $\hat{\xi}_u$ plotted against the true value, ξ, for the Bleasdale-Nelder data set. The 45° line represents zero error. The prior mean (x_m) is 99.0 and this is marked on the figure as a horizontal dotted line. Fig. 4.2 shows that $\hat{\xi}_{Tr}$ can be thought of as a shift of $\hat{\xi}_u$ towards the prior mean $\bar{x}$, the shift being largest when the calibration data are least informative. For the Bleasdale-Nelder calibration data, as ξ or X increases the spline smoother becomes flatter (see Fig. 3.10). This is reflected generally in larger shifts of $\hat{\xi}_u$ towards the prior mean for $\xi > 140$ approx. (see Fig. 4.2).

4.6.3 A non-central Student prior density

Suppose the calibration experiment consists of N observations (x_i, Y_i) $i=1,2.....N$, $\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}$ and $S_{xx} = \sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{(N-1)}$. A truncated Student prior density was assumed for ξ. It is defined by

$$\pi(\xi) \propto \frac{1}{[(1+\frac{1}{N})S_{xx}]^{\frac{1}{2}}\,[1+\frac{(\xi-\bar{x})^2}{(1+\frac{1}{N})S_{xx}}]^{N/2}} \qquad x_1 \le \xi \le x_n$$

$$= 0 \qquad \text{otherwise} \qquad (4.6.3)$$

This is the density function of a Student distribution based on $(N-1)$ degrees of freedom, centred at $\bar{x}$ and with scale factor $(1+\frac{1}{N})\,\frac{S_{xx}}{(N-1)}$, which has been truncated at $\xi = x_1$ and $\xi = x_n$. The maximum of $\pi(\xi)$ occurs at $\xi = \bar{x}$ so π_{MAX} is given by

$$\pi_{\text{MAX}} = \frac{C}{[(1+\frac{1}{N})S_{xx}]^{\frac{1}{2}}}$$

where C is a normalising constant. Suppose the posterior density of ξ corresponding to this Student density is called $\pi_t(\xi|\mathbf{Y}')$ and the corresponding posterior density for ζ is called $\pi_t(\zeta|\mathbf{Y}')$. In order to simulate from $\pi(\xi|\mathbf{Y}')$ we must first simulate from $\pi_t(\zeta|\mathbf{Y}')$ using the following algorithm:

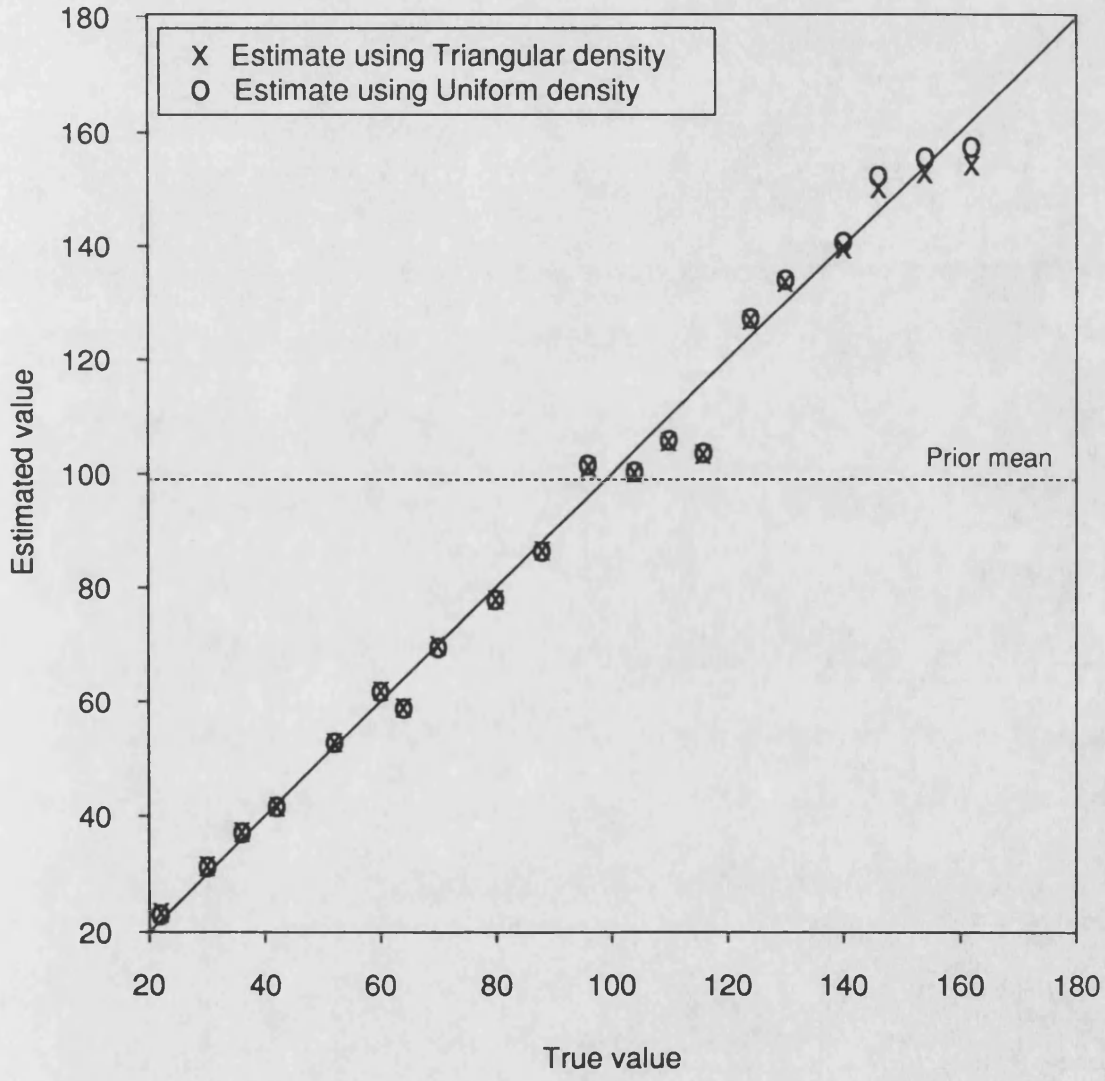
(a) Simulate ζ^* from $\pi_u(\zeta|\mathbf{Y}')$

(b) Accept ζ^* iff

$$U\,\pi_{\text{MAX}} \le \pi(\tilde{f}^{-1}(\frac{\zeta^*-b}{a}))$$

Fig.4.2: Comparison of point estimates

(Bleasdale-Nelder data set)



i.e. if and only if

$$U \leq \left[1+\frac{(\xi^*-\bar{x})^2}{(1+\frac{1}{N})S_{xx}}\right]^{-N/2} \qquad (4.6.4)$$

where $\xi^* = \tilde{f}^{-1}\left(\frac{\zeta^*-b}{a}\right)$ and U is an independent $U(0,1)$ random variable.

Using the algorithm (4.6.4) 50 realisations of $\pi_t(\zeta|\mathbf{Y}')$ were obtained. Suppose these are denoted by ζ_v^* (v=1,2,...,50). For each posterior realisation of f, $\tilde{f}_s$, the values of

$$\tilde{f}_s^{-1}\left(\frac{\zeta_v^*-b}{a}\right) \qquad \begin{matrix} s=1,2,...,199,\ v=1,2,...,50 \\ s=200,\ v=1,2,3....49 \end{matrix}$$

were calculated giving 9999 realisations of $\pi_t(\xi|\mathbf{Y}')$. By way of illustration, Table 4.5 shows the mean absolute errors of point estimates of ξ for the Weibull, Asymptotic and Linear data sets. Table 4.6 shows the maximum width, minimum width and mean width of the twenty interval estimates of ξ together with the proportion of successful estimates. The probability level for the interval estimates is 90%.

Let $\hat{\xi}_t$ be the point estimate of ξ when one assumes a non-central Student prior density for ξ and let $\hat{\xi}_u$ be the point estimate of ξ when one assumes a uniform density for ξ. Fig. 4.3 shows $\hat{\xi}_t$ and $\hat{\xi}_u$ plotted against the true value, ξ, for the Asymptotic data set. The 45° line represents zero error. The prior mean $\bar{x}$ is 11.27 and this is marked on the figure as a horizontal dotted line. Fig. 4.3 shows that $\hat{\xi}_t$ can be thought of as a shift of $\hat{\xi}_u$ towards this prior mean , the shift being largest when the calibration data are least informative. The Asymptotic calibration data are least informative for larger values of ξ (see Fig. 3.11). As ξ or X increases, the spline smoother becomes flatter and approaches the horizontal. This is reflected generally in larger shifts of $\hat{\xi}_u$ towards the prior mean for the larger values of ξ (approx $\xi > 11$). There are also larger errors in both estimates, $\hat{\xi}_u$ and $\hat{\xi}_t$, for the larger values of ξ (see Fig. 4.3) which is a reassuring feature of the non-parametric method because the data are not as informative concerning the value of ξ in the region where the spline is flat.

A comparison of Tables 4.2, 4.4 and 4.6 shows that as expected, when an informative prior such as a triangular or Student prior is used, the mean interval width is reduced . However it will be noted that the reductions are in general small. This is because for all the calibration data sets the number of observations, N, is 240. This means that the calibration data are very informative concerning the value of ξ and the particular choice of prior will not greatly affect the posterior

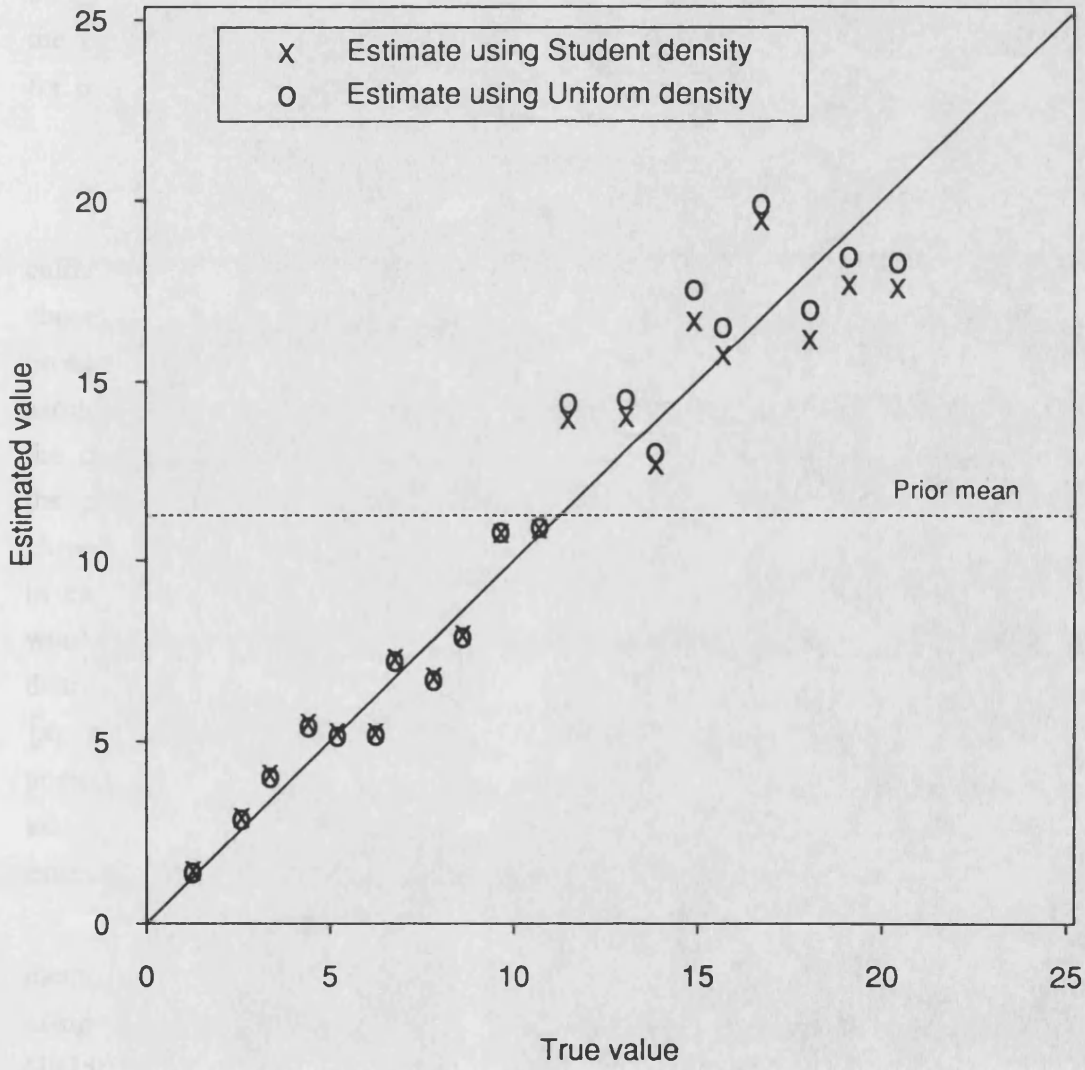
Table 4.5 Point estimates for ξ assuming a non-central Student prior for ξ

Data set	Calibration range	Mean absolute error
Weibull	10.00 - 79.52	0.72
Asymptotic	1.00 - 21.74	1.12
Linear	15.00 - 70.00	3.32

Table 4.6 Interval estimates for ξ assuming a non-central Student prior for ξ (90% probability level)

Data set	Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Weibull	7.88	1.86	3.52	19/20
Asymptotic	7.71	1.23	4.84	19/20
Linear	16.22	10.93	14.97	18/20

Fig.4.3: Comparison of point estimates
(Asymptotic data set)



distribution of ξ. The parametric methods of Hoadley (1970) and Aitchison and Dunsmore (1975) are applied to the Linear data set in Chapter 7 using the same uniform prior as used here. The results obtained using the non-parametric method described in this chapter compare very favourably with these. A comparison of all the results is made in Chapter 7.

4.7 DATA ON LENGTH OF TRANSPARENT DENTINE

Data from upper central, upper canine, lower central, lower lateral and lower canine teeth were considered. The same data sets were considered in Section 3.3.2, Chapter 3. For these data sets the total numbers of observations, (N), in the calibration experiments are 31, 15, 28, 24 and 28 respectively. The model for the prediction stage of the calibration process is given by

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3 \text{ or } 4$$

The calibration experiments involving the teeth data are not controlled calibration experiments because the ages of patients were not selected or pre-chosen. The patients presented themselves at the hospital with a variety of problems and their teeth were extracted. It might therefore be argued that if this sample of patients is typical of future samples of patients presenting themselves at the dental hospital, then the calibration experiments provide some indication of the pattern of ages and can therefore be used to specify a prior density for ξ. However there are only 43 patients in the sample and the number of observations in each of the five calibration experiments is quite small. It was decided that it would be better therefore to assume a non-informative prior for ξ. Posterior distributions $\pi(\xi|\mathbf{Y}')$ were constructed assuming a uniform prior density on $[x_1, x_n]$ for ξ. The procedure was the same as that given in Section 4.6.1. These posterior distributions were for new patients whose ages were *known* to enable assessment of the method. Table 4.7 gives details of the point and interval estimates obtained. The probability level for the interval estimates is 80%.

All the results obtained for the five types of teeth using the non-parametric method described in this chapter compare very favourably with those obtained using the parametric approaches of Hoadley (1970) and Aitchison and Dunsmore (1975). A comparison of these results is given in Chapter 7.

4.8 ANTIBIOTIC ASSAY DATA

The methods discussed in Sections 4.2 - 4.4 were applied to the antibiotic assay data first considered in Section 3.3.3, Chapter 3. There was no prior information about ξ so a uniform prior density for ξ was assumed.

Table 4.7 Teeth data sets - estimation results assuming a uniform prior for ξ (80% probability level)

Type of tooth	Upper central	Lower central	Upper canine	Lower lateral	Lower canine
Mean absolute error (yrs)	7.7	5.6	4.7	5.7	9.8
Maximum width of interval estimates (yrs)	23.6	23.5	20.8	24.0	25.0
Minimum width of interval estimates (yrs)	12.8	9.3	11.0	13.0	10.1
Mean width of interval estimates (yrs)	17.7	14.0	15.0	20.2	19.5
No. of successful estimates	10	2	5	5	9
Total no. of new teeth	15	3	7	6	13

The model for the prediction stage of the calibration process is given by

$$Y_j' = f(\xi) + \varepsilon_j' \quad j=1.$$

Posterior distributions $\pi(\xi|Y')$ were obtained using the methods described above. Table 4.8 gives details of the mean absolute errors and interval estimates for ξ. The probability level for the interval estimates is 90%.

4.9 CONCLUDING REMARKS

The method considered in this chapter used simulation and a switching algorithm to generate the posterior distribution of ξ. Unlike the previous chapter, which assumes a prior for η and hence an implicit prior for ξ, the method described in this chapter enables one to use an explicit prior for ξ. Three different prior densities for ξ were used to illustrate the versatility of the method, a uniform prior on $[x_1, x_n]$, a triangular prior on $[x_1, x_n]$ and a truncated Student prior. The method worked well on all the data sets and consistently produced good point and interval estimates for ξ. The posterior median was used as a point estimator of ξ and symmetrical interval estimates were constructed from the posterior distributions of ξ.

Any prior density defined on the calibration range $[x_1, x_n]$ can be used provided it is unimodal within $[x_1, x_n]$, so a wide variety of prior information on ξ can be easily incorporated into the calibration process. The method is easy to implement and is computationally efficient. There are no obvious disadvantages with the method.

Table 4.8. Antibiotic assay data set - estimation results assuming a uniform prior for ξ (90% probability level)

Mean absolute error	0.069
Maximum width of interval estimates	0.43
Minimum width of interval estimates	0.12
Mean width of interval estimates	0.30
No. of successful estimates	57
Total no. of estimates	60
Calibration range	-0.22 to 1.00

5. AN APPROACH USING PREDICTIVE DENSITIES

5.1 INTRODUCTION

In the previous two chapters, 3 and 4, ξ was viewed as a non-linear functional of the true calibration curve, f, and simulation was used to generate the posterior distribution of ξ. In this chapter a completely different non-parametric approach is considered which does not use simulation but which is based on predictive densities. It has certain similarities with the parametric Bayesian approaches of Hoadley (1970) and Aitchison and Dunsmore (1975). They used a predictive density approach when considering the calibration process under a linear model. The models for the calibration experiment and prediction stage of the calibration process respectively are the same as those used in Chapters 3 and 4, namely equations (3.2.1) and (3.2.2). Again $\hat{f}$ denotes the spline smoother of the calibration data set.

Suppose that we have observations $Y_1, Y_2, \ldots, Y_n$ arising from the model

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,n \tag{5.1.1}$$

where f is a natural cubic smoothing spline with knots $\{x_i\}$ and $a < x_1 < x_2, \ldots, < x_n < b$. Also assume that the errors ε_i are independently identically distributed normal random variables with mean zero and variance of σ^2. The whole approach of this chapter hinges on the approximation of the distribution of the residual sum of squares. The residual sum of squares (RSS) for the above model is given by

$$RSS = (\mathbf{Y}-\hat{\mathbf{f}})^{\mathrm{T}}(\mathbf{Y}-\hat{\mathbf{f}}) \tag{5.1.2}$$

where $\mathbf{Y}^{\mathrm{T}} = (Y_1\, Y_2 \ldots\ldots Y_n)$ and $\hat{\mathbf{f}}^{\mathrm{T}} = (\hat{f}(x_1)\, \hat{f}(x_2) \ldots\ldots \hat{f}(x_n))$. If one lets the smoothing parameter α tend to infinity, $\hat{f}$ tends to the least squares regression line. In the case of a linear model, i.e. $f(x) = \beta_0 + \beta_1 x$

$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-2}$$

where χ^2_υ is a chi-squared random variable based on υ degrees of freedom. However, in the case of f being a natural cubic smoothing spline, the distribution of RSS is not obtainable in any closed analytical form and must be approximated.

It will be recalled from Section 3.1.3, Chapter 3, that

$$\hat{f} = A(\alpha)\mathbf{Y}$$

where $A(\alpha)$ is the *hat matrix*. The matrix $A(\alpha)$ is symmetric and positive definite. Substituting for $\hat{f}$ in equation (5.1.2) gives

$$\text{RSS} = \mathbf{Y}^{\mathrm{T}}(I-A)^2\mathbf{Y}.$$

Let $D = (I-A)^2$ then D is symmetric and positive definite and

$$\text{RSS} = \mathbf{Y}^{\mathrm{T}}D\mathbf{Y} \tag{5.1.3}$$

Hence the residual sum of squares is a quadratic form in the observations $Y_1, Y_2, \ldots, Y_n$. Using the fact that

$$\mathbf{Y} = \mathbf{f} + \boldsymbol{\varepsilon} \text{ where } \boldsymbol{\varepsilon}^{\mathrm{T}} = (\varepsilon_1\, \varepsilon_2, \ldots, \varepsilon_n) \text{ and } \mathbf{f}^{\mathrm{T}} = (f(x_1)\, f(x_2) \ldots f(x_n))$$

$$\begin{aligned}\text{RSS} &= (\mathbf{f}+\boldsymbol{\varepsilon})^{\mathrm{T}} D(\mathbf{f}+\boldsymbol{\varepsilon}) \\ &= \mathbf{f}^{\mathrm{T}}D\mathbf{f} + 2\mathbf{f}^{\mathrm{T}}D\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^{\mathrm{T}}D\boldsymbol{\varepsilon}, \end{aligned} \tag{5.1.4}$$

5.1.1 The bias term

The first term on the right-hand side of equation (5.1.4), $\mathbf{f}^{\mathrm{T}}D\mathbf{f}$, is known as the *bias term*. This arises because the bias is defined as follows:

$$\text{Bias} = \mathbf{f} - E(\hat{\mathbf{f}}) = \mathbf{f} - E(A\mathbf{Y}) = (I-A)\mathbf{f}$$

So $\mathbf{f}^{\mathrm{T}}D\mathbf{f} = \mathbf{f}^{\mathrm{T}}(I-A)^2\mathbf{f}$ is the sum of the squares of the biases.

Assume that the knots x_i $(i=1,2,\ldots,n)$ are generated by a density $p(x)$ on $[a,b]$ where

$$\int_a^{x_j} p(x)dx = (2j-1)/2n$$

Assume without loss of generality that $[a,b] = [0,1]$. Define the integrated squared bias to be

$$B_n^2(\lambda) = \int_0^1 [f(x) - E(\hat{f}(x))]^2\, p(x)dx$$

where λ is a multiple of the smoothing parameter α. In particular $\lambda = \alpha/n$.

For large n ,

$$\begin{aligned}\mathbf{f}^{\mathrm{T}}D\mathbf{f} &= \sum_{i=1}^{n} [f(x_i) - E(\hat{f}(x_i))]^2 \\ &\simeq n\int_0^1 [f(x) - E(f(x))]^2\, p(x)dx \\ &= n\, B_n^2(\lambda).\end{aligned}$$

Let $W_m[0,1]$ denote the space of functions f on $[0,1]$ such that $D^j f$, $j \le (m-1)$ is absolutely continuous and $D^m f \in \mathcal{L}_2$ where $\mathcal{L}_2$ is the set of measurable square

integrable functions on $[0,1]$ and D represents the differentiation operator. Eubank (1988) states that for any $f \in W_m[0,1]$ and for large n, $B_n^2(\lambda)$ is bounded above by $\lambda J_m(f)$ where $J_m(f)$ is defined as $\int_0^1 D^m f(x)\,dx$. If one defines the integrated risk $IR_n(\lambda)$ to be

$$\int_0^1 E[f(x) - \hat{f}(x)]^2 p(x)\,dx$$

then for large n, $IR_n(\lambda)$ is bounded above by

$$\lambda J_m(f) + \frac{\sigma^2 l_m}{n\lambda^{\frac{1}{2m}}}\,\pi^{-1}\int_0^1 p(x)^{\frac{1}{2m}}\,dx$$

where $l_m = \int_{-\infty}^{\infty}(1+y^{2m})^{-2}\,dy$ (Eubank 1988).

Hence if one takes $\lambda = 0(n^{-2m/2m+1})$ then $IR_n(\lambda)$ can always be made to decay at a rate of $n^{-2m/(2m+1)}$ which is the best uniform rate of convergence for functions in $W_m[0,1]$. Speckman (1989) shows that if $f \in W_{2m}[0,1]$ and if in addition

$$f^{(j)}(0) = f^{(j)}(1) = 0 \qquad j = m,(m+1),\ldots,(2m-1)$$

and $p(x) \equiv 1$ then $B_n^2(\lambda)$ is bounded above by $\lambda^2 J_{2m}(f)$. In this case the optimal λ is $0(n^{-2m/(4m+1)})$. In this thesis we have assumed f is a *natural* spline which means that the above conditions hold. Also we are interested in the case of $m = 2$. So for large n, $\mathbf{f}^T D\mathbf{f} \sim n\lambda^2 J_4(f)$ and $\mathbf{f}^T D\mathbf{f}$ is $0(n^{1/9})$.

5.1.2 The term $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$

The term $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$ is a quadratic form in normal random variables which has mean $\sigma^2 \operatorname{tr} D$ and variance $2\sigma^4 \operatorname{tr} D^2$.

$$E(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}) = \sigma^2 \operatorname{tr} D = \sigma^2(n - tr A)$$

where A is the *hat matrix*. Let $T_n(\lambda)$ be defined as follows:

$$T_n(\lambda) = n^{-1}\operatorname{trace} A(\lambda) = n^{-1}\sum_{j=1}^{n} A_{jj}(\lambda).$$

Silverman (1984) showed that as $n \to \infty$

$$n\lambda^{1/4} T_n(\alpha) \to 2^{-3/2}\int_a^b [f(x)]^{1/4}\,dx.$$

As stated in the previous section, the optimal λ is $0(n^{-2m/4m+1})$, i.e. $0(n^{-4/9})$ so for large n, $tr(A)$ is $0(n^{1/9})$. Hence for large n, $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$ is $0(n)$.

5.1.3 The random perturbation term

The middle term of the right-hand side of equation (5.1.4) $2\mathbf{f}^T D\boldsymbol{\varepsilon}$, is a linear combination of independent normal random variables so it is normally distributed with mean zero and variance $4\sigma^2\mathbf{f}^T D^2\mathbf{f}$. It is known as the random perturbation term. Let $\mathbf{x} = (I-A)\boldsymbol{\varepsilon}$ and let $\mathbf{y} = (I-A)\mathbf{f}$. The term $2\mathbf{f}^T D\boldsymbol{\varepsilon}$ is therefore $2(\mathbf{x} \bullet \mathbf{y})$ where $\mathbf{x} \bullet \mathbf{y}$ is the scalar product of two vectors. By Schwartz's inequality

$$(\mathbf{x} \bullet \mathbf{y})^2 \leq \|\mathbf{x}\|^2 \, \|\mathbf{y}\|^2$$

Applying this inequality here gives

$$(\mathbf{f}^T D\boldsymbol{\varepsilon})^2 \leq (\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon})\,(\mathbf{f}^T D\mathbf{f})$$

Hence for large n, the middle term $2\mathbf{f}^T D\boldsymbol{\varepsilon}$ is at most $0(n^{5/9})$.

5.1.4 The distribution of RSS

The results given in Sections 5.1.1, 5.1.2 and 5.1.3 above show that of the three terms on the right-hand side of equation (5.1.4) the dominant term is $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$. Also $E(RSS) = \mathbf{f}^T D\mathbf{f} + E(\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon})$ and $E(RSS-\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}) = \mathbf{f}^T D\mathbf{f}$. It is argued therefore that provided n is reasonably large, the distribution of RSS, which is itself a quadratic form in $\mathbf{Y}$, will be close to the distribution of $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$, a quadratic form in normal random variables. Hence to derive an approximation to the distribution of RSS we are led to studying approximations to quadratic forms in normal random variables.

Quadratic forms in normal variables can be represented as weighted sums of independent chi-squared random variables (Johnson and Kotz, 1970). In particular, let Q_n be a quadratic form defined as

$$Q_n = \mathbf{Z}^T B\mathbf{Z}$$

where $\mathbf{Z}^T = (Z_1, Z_2, \ldots, Z_n)$ and $\mathbf{Z}$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance matrix Σ. Suppose B is an $n \times n$ symmetric matrix. The distribution of Q_n is the same as that of $\sum_{i=1}^{n} \lambda_i W_i^2$ where W_i are independently distributed $N(0,1)$ random variables $(i=1,2,\ldots n)$ and $\lambda_1, \lambda_2, \lambda_3 \ldots\ldots \lambda_n$ are the eigenvalues of ΣB $(\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \geq \lambda_n)$. Here we are considering the distribution of $Q_n = \boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon}$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance matrix $\sigma^2 I_n$. So $\lambda_1, \lambda_2, \lambda_3 \ldots\ldots \lambda_n$ are the eigenvalues of the matrix $\sigma^2 D$. Since $\sigma^2 > 0$ and D is positive definite $\lambda_i > 0$ $(i=1,2,\ldots,n)$. Also one can state that

$$E(Q_n) = \sum_{k=1}^{n} \lambda_k = \sigma^2 \operatorname{tr} D$$

$$\operatorname{Var}(Q_n) = 2 \sum_{k=1}^{n} \lambda_k^2 = 2\sigma^4 \operatorname{tr} D^2 \qquad (5.1.5)$$

$$E(Q_n - E(Q_n))^3 = 8 \sum_{k=1}^{n} \lambda_k^3 = 8\sigma^6 \operatorname{tr} D^3$$

The distribution function or density function of such weighted sums of chi-squared variables can be computed almost exactly by numerically inverting the characteristic function (Imhof, 1961). This approach will be considered in Section 5.4. Alternatively there are three approximations suggested in the literature. The best known approximation is that of Welch and Satterthwaite which approximates the distribution of a quadratic form by that of $c\chi_v^2$ with c and v chosen to equate the first two moments of the two distributions (Fairfield-Smith, 1936; Satterthwaite, 1941, 1946; Welch, 1936, 1937). The second approximation is due to Solomon and Stephens (1977) and uses $c(\chi_p^2)^d$ with the parameters being chosen to equate the first three moments of the two distributions. The equations defining c, d and p require an iterative solution. The third approximation is that of Eagleson and Buckley (1988) which approximates the distribution of Q_n by $c\chi_v^2 + d$. This approximation was fitted using the first three moments and is quite easy to apply requiring only the evaluation of traces of matrices. The first two approximations have no theoretical justification whereas that of Eagleson-Buckley is based on the Edgeworth expansion. Eagleson and Buckley found that in the cases of non-parametric regression which they considered, their approximation was as expected better than the Welch-Satterthwaite approximation but not as good as the Solomon-Stephens approximation. It was therefore decided to consider the approximations of Solomon-Stephens and Eagleson-Buckley only.

5.2 A PREDICTIVE DENSITY APPROACH

Suppose we specify a density, $\pi(\theta)$, called the prior density of θ, on a space Θ. The first stage of the calibration process is to perform an informative experiment called the calibration experiment. Suppose $\mathcal{S}$ is the sample space associated with this calibration experiment and the calibration data are denoted by $\mathbf{z}$ where

$$\mathbf{z} = \{(x_1, Y_1), (x_2, Y_2), \ldots, (x_N, Y_N)\}$$

The calibration experiment is described by a set of probability distributions on $\mathcal{S}$, namely

$$\{p(\mathbf{z}|\theta) : \theta \in \Theta\} \qquad \mathbf{z} \in \mathcal{S}$$

The information obtained by observing $\mathbf{z}$ in the calibration experiment clearly influences the prior density $\pi(\theta)$ and leads to a posterior density for θ, $\pi(\theta|\mathbf{z})$ where

$$\pi(\theta|\mathbf{z}) = \pi(\theta)\frac{p(\mathbf{z}|\theta)}{\pi(\mathbf{z})}$$

and $\pi(\mathbf{z}) = \int_{\Theta} \pi(\theta)p(\mathbf{z}|\theta)\,d\theta$.

The second stage of the calibration process is a prediction stage and involves a future experiment where the observations $\mathbf{Y}'$ are made. There is uncertainty about which density $p(\mathbf{Y}'|\theta)$ applies to this future experiment but an assessment of $\pi(\theta|\mathbf{z})$ has been made. A essential part of the Bayesian approach to prediction is the probability distribution of a future observation $\mathbf{Y}'$ given the outcome $\mathbf{z}$ of the informative experiment. This is defined by

$$\begin{aligned} p(\mathbf{Y}'|\mathbf{z}) &= \int_{\Theta} p(\mathbf{Y}'|\theta)\,\pi(\theta|\mathbf{z})\,d\theta \\ &\propto \int_{\Theta} p(\mathbf{Y}'|\theta)\,p(\mathbf{z}|\theta)\pi(\theta)d\theta. \end{aligned} \tag{5.2.1}$$

The density function $p(\mathbf{Y}'|\mathbf{z})$ is called the *predictive density function* for $\mathbf{Y}'$ given $\pi(\theta)$ and $\mathbf{z}$. Suppose the observation $\mathbf{Y}'$ in the future experiment corresponds to an unknown value of X, namely ξ. The aim of the calibration process is to derive the posterior density of ξ given $\mathbf{Y}'$ and $\mathbf{z}$ and predictive densities are an essential part of this derivation.

Let $\mathbf{z} = (\mathbf{x}, \mathbf{Y})$ where $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, x_3, \ldots, x_N)$ and $\mathbf{Y}^{\mathrm{T}} = (Y_1, Y_2, \ldots, Y_N)$. Aitchison and Dunsmore (1975) showed that for a random calibration experiment

$$\pi(\xi|\mathbf{Y}', \mathbf{z}) \propto \pi(\xi|\mathbf{x})p(\mathbf{Y}'|\xi, \mathbf{z}) \tag{5.2.2}$$

and for a controlled calibration experiment

$$\pi(\xi|\mathbf{Y}', \mathbf{z}) \propto \pi(\xi)p(\mathbf{Y}'|\xi, \mathbf{z}) \tag{5.2.3}$$

where $p(\mathbf{Y}'|\xi, \mathbf{z})$ is the predictive density associated with the future experiment and based on data $\mathbf{z}$. The prior beliefs about ξ are contained in the densities $\pi(\xi|\mathbf{x})$ and $\pi(\xi)$. The argument is that in a random calibration experiment where the observations $x_1, x_2, \ldots, x_N$ occur randomly, any prior beliefs about ξ will depend on the pattern observed in $x_1, x_2, \ldots, x_N$ whereas in a controlled calibration experiment, the values $x_1, x_2, \ldots x_N$ are selected by the experimenter so the calibration experiment provides no direct information concerning the plausibilities of various ξ. Aitchison and Dunsmore comment that expressions (5.2.2) and (5.2.3) take the form of Bayes theorem, as it could be applied after the data $\mathbf{z}$ or $(\mathbf{x}, \mathbf{Y})$ of the calibration experiment are known. The prior density is $\pi(\xi|\mathbf{x})$ or $\pi(\xi)$ and the *likelihood* of Bayes theorem takes the form of the predictive

distribution. In fact Hoadley (1970) calls the term $p(\mathbf{Y}'|\mathbf{z},\xi)$, $L(\xi)$, saying that it is a kind of likelihood function representing the information on ξ from all sources except that associated with the prior density of ξ.

Consider the model (5.1.1) as the model associated with the calibration experiment

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,N$$

i.e. suppose, for simplicity of argument, that $n = N$ and all the x_i are distinct. Suppose the residual sum of squares associated with the calibration experiment is defined as

$$RSS(\alpha) = \sum_{i=1}^{N} (Y_i - \hat{f}(x_i))^2$$

where $\hat{f}$ is the spline smoother for the calibration data and α is the smoothing parameter. Suppose that the model for the prediction stage of the calibration process, i.e. the future experiment, is given by

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,\ldots,m \tag{5.2.4}$$

where it is assumed that the ε_j' are i.i.d. normal random variables with mean zero and variance σ^2. They are also assumed to be independent of the ε_i. Let $\bar{Y}' = \sum_{j=1}^{m} Y_j'/m$ and suppose V_f represents the residual sum of squares associated with the future experiment, i.e.

$$V_f = \sum_{j=1}^{m} (Y_j' - \bar{Y}')^2. \tag{5.2.5}$$

It will be recalled from equation (3.1.12) that f can be written as a linear combination of n B-splines ($\beta_i(x)$), so that

$$f(x) = \sum_{i=1}^{n} \gamma_i \beta_i(x).$$

Let $\gamma^{\mathrm{T}} = (\gamma_1 \gamma_2 \ldots\ldots \gamma_n)$. The likelihood associated with models (5.1.1) and (5.2.4) is

$$\begin{aligned}
&p(\mathbf{z},Y_1',Y_2',\ldots,Y_m' \mid \sigma^2,\xi,\gamma) \\
&\propto p(\mathbf{z} \mid \sigma^2,\gamma)\, p(\mathbf{Y}' \mid \sigma^2,\xi,\gamma) \\
&= p(\mathbf{z} \mid \sigma^2,\gamma) \frac{1}{\sigma^m} \exp\left[- \sum_{j=1}^{m} \frac{(Y_j' - f(\xi))^2}{2\sigma^2} \right] \qquad \text{where } f(\xi) = \sum_{i=1}^{n} \gamma_i \beta_i(\xi)
\end{aligned}$$

$$= p(\mathbf{z} \mid \sigma^2,\gamma)\ \frac{1}{\sigma^m} \exp\left[- \frac{\sum_{j=1}^{m}(Y_j'-\bar{Y}')^2 + m(\bar{Y}'-f(\xi))^2}{2\sigma^2} \right].$$

Hence $(\mathbf{z},V_f,\bar{Y}')$ is a sufficient statistic and

$$\pi(\xi \mid \mathbf{z},\mathbf{Y}') = \pi(\xi \mid \mathbf{z},V_f,\bar{Y}')$$

By Bayes theorem

$$\begin{aligned}\pi(\xi \mid \mathbf{z},\mathbf{Y}') &\propto \pi(\xi \mid \mathbf{z})p(V_f,\bar{Y}' \mid \xi,\mathbf{z}) \\ &= \pi(\xi \mid \mathbf{z})p(\bar{Y}' \mid \xi,\mathbf{z},V_f)p(V_f \mid \xi,\mathbf{z}) \\ &\propto \pi(\xi \mid \mathbf{z})p(\bar{Y}' \mid \xi,\mathbf{z},V_f)\end{aligned}$$

since $p(V_f \mid \mathbf{z}, \xi)$ does not depend on ξ. So for a random calibration experiment

$$\pi(\xi \mid \mathbf{z},\mathbf{Y}') \propto \pi(\xi \mid \mathbf{x})p(\bar{Y}' \mid \xi,\mathbf{z},V_f)$$

and for a controlled calibration experiment

$$\pi(\xi \mid \mathbf{z},\mathbf{Y}') \propto \pi(\xi)p(\bar{Y}' \mid \xi,\mathbf{z},V_f)$$

One can see from the above expressions for the posterior density of ξ, that the particular predictive density of importance here is the predictive density of $\bar{Y}'$, $p(\bar{Y}' \mid \xi,\mathbf{z},V_f)$.

Let $L(\xi) = p(\bar{Y}' \mid \xi,\mathbf{z},V_f)$ then

$$\pi(\xi \mid \mathbf{z},\mathbf{Y}') \propto \begin{matrix}\pi(\xi \mid \mathbf{x})L(\xi) \\ \pi(\xi)L(\xi)\end{matrix} \qquad (5.2.6)$$

Before obtaining the posterior density of ξ it is necessary to prove a few results.

Theorem 5.1. Suppose that *a priori*, ξ is independent of (γ,σ^2) and the prior distribution of γ is given by

$$p(\gamma \mid \sigma^2) \ \propto\ -\ \frac{\alpha}{2\sigma^2} \int (f''(x))^2 dx = -\frac{\alpha}{2\sigma^2}\, \gamma^{\mathrm{T}}\, \Omega\, \gamma$$

where Ω is an $n \times n$ symmetric banded matrix defined by

$$\Omega_{ij} = \int_{-\infty}^{\infty} \beta''_i(x)\, \beta''_j(x) dx$$

and β_i is the ith B-spline. Then

$$\{\bar{Y}' \mid \sigma^2,\xi,\mathbf{z},V_f\} \sim N(\, \beta^{\mathrm{T}}\hat{\gamma},\ (\beta^{\mathrm{T}} S^{-1} \beta + \frac{\sigma^2}{m}))$$

where

$\beta^T = (\beta_1(\xi)\, \beta_2(\xi),\ldots,\beta_n(\xi))$, $S = \frac{1}{\sigma^2}(B^TB + \alpha\Omega)$ and $\hat{\gamma} = \frac{1}{\sigma^2}S^{-1}B^T\mathbf{Y}$

(see Chapter 3, Sections 3.1.3 and 3.2.2 for further information on Ω, β_i and S).

Proof. Conditional on $(\sigma^2, \mathbf{z})$, (ξ, V_f) and γ are independent, i.e.

$$p(\gamma \mid \sigma^2, \mathbf{z}, \xi, V_f) = p(\gamma \mid \sigma^2, \mathbf{z})$$

Silverman (1985) showed that, with the given prior $-\frac{\alpha}{2\sigma^2}\gamma^T\Omega\gamma$, the posterior distribution of γ is a multivariate normal distribution with mean $\hat{\gamma}$ and variance matrix S^{-1} where $\hat{\gamma} = (B^TB + \alpha\Omega)^{-1}B^T\mathbf{Y}$ and $S^{-1} = (B^TB + \alpha\Omega)^{-1}\sigma^2$. We shall write

$$\{\gamma \mid \sigma^2, \mathbf{z}, \xi, V_f\} \sim \mathcal{N}_n(\hat{\gamma}, S^{-1})$$

Now from the assumptions of model (5.2.4)

$$\{\bar{\varepsilon}' \mid \sigma^2, \xi, V_f, \gamma\} \sim N(0, \frac{\sigma^2}{m})$$

where $\bar{\varepsilon}' = \sum_{j=1}^{m} \varepsilon_j' / m$. So the joint density of γ and $\bar{\varepsilon}'$ is given by

$$(\gamma, \bar{\varepsilon}') \sim \mathcal{N}_{n+1}(\mu, \Sigma)$$

where $\mu^T = (\hat{\gamma}_1\, \hat{\gamma}_2\, \hat{\gamma}_3\, \ldots..\, \hat{\gamma}_n\, 0)$ and Σ is an $(n+1)$ by $(n+1)$ partitioned matrix given by

$$\sum = \sigma^2 \begin{bmatrix} S_1^{-1} & 0_{n\times 1} \\ 0_{1\times n} & 1/m \end{bmatrix}.$$

Here $S_1 = \sigma^2 S = (\alpha\Omega + B^TB)$. Let β be an $n \times 1$ column vector of B-splines evaluated at the point $X = \xi$. From model (5.2.4)

$$\begin{aligned} \bar{Y}' &= f(\xi) + \bar{\varepsilon}' \\ &= \beta_1(\xi)\gamma_1 + \beta_2(\xi)\gamma_2 + \ldots + \beta_n(\xi)\gamma_n + \bar{\varepsilon}' \\ &= C\gamma^* \end{aligned}$$

where $C = (\beta^T\ 1)$ and $\gamma^{*T} = (\gamma_1\, \gamma_2 \ldots..\, \gamma_n\, \bar{\varepsilon}')$, so the distribution of $\bar{Y}'$ is normal with mean $C^T\mu$ and variance $C\sum C^T$.

In particular

$$C\mu = \beta_1(\xi)\hat{\gamma}_1 + \beta_2(\xi)\hat{\gamma}_2 + \beta_3(\xi)\hat{\gamma}_3 + ,\ldots,+ \beta_n(\xi)\hat{\gamma}_n + 1\text{x}0$$

$$= \beta^T \hat{\gamma}$$

and $C \sum C^T$ is given by

$$(\beta_1(\xi)\ \beta_2(\xi) \ldots\ldots \beta_n(\xi)\ 1) \begin{bmatrix} \sigma^2 S_1^{-1} & 0_{n\times 1} \\ 0_{1\times n} & \sigma^2/m \end{bmatrix} \begin{bmatrix} \beta_1(\xi) \\ \beta_2(\xi) \\ \\ \beta_n(\xi) \\ 1 \end{bmatrix}$$

So $C \sum C^T = \beta^T S^{-1} \beta + \sigma^2/m$ where $S = \frac{1}{\sigma^2}(\alpha\Omega + B^T B)$

Corollary 5.1. For the case of the calibration model being weighted, i.e. the observations Y_i have variances given by $\mathrm{Var}(Y_i) = \sigma^2/w_i$ $(i=1,2,\ldots,n)$ the results become

$$\{\bar{Y}' \mid \sigma^2, \xi, \mathbf{z}, V_f\} \sim N(\beta^T \hat{\gamma},\ (\beta^T S^{-1} \beta + \frac{\sigma^2}{m}))$$

where β^T is as defined in Theorem 5.1 and $\hat{\gamma}$, S are given given by

$$\hat{\gamma} = \frac{1}{\sigma^2} S^{-1} B^T W \mathbf{Y} \quad \text{and} \quad S = \frac{1}{\sigma^2}(B^T W B + \alpha\Omega).$$

Here W is the $n \times n$ diagonal matrix with entries w_i and $\mathbf{Y}$ is an $(n \times 1)$ vector of mean values where $\mathbf{Y}^T = (\bar{Y}_1\ \bar{Y}_2, \ldots, \bar{Y}_n)$ and $\bar{Y}_i$ is the mean of the observations at knot i.

Let $\tau = 1/\sigma^2$. If one uses result (5.2.1) with $\theta = \tau$ then the predictive density of $\bar{Y}'$ given $\mathbf{z}$, ξ and V_f is given by

$$L(\xi) \propto \int_0^\infty p(\bar{Y}' \mid \tau, \xi, \mathbf{z}, V_f)\, \pi(\tau \mid \xi, \mathbf{z}, V_f)\, d\tau \qquad (5.2.7)$$

Using Bayes theorem

$$\pi(\tau \mid \xi, \mathbf{z}, V_f) \propto \pi(\tau \mid \xi, \mathbf{z})\, p(V_f \mid \tau, \xi, \mathbf{z})$$

Now V_f is the residual sum of squares at the second stage of the calibration process and is given by equation (5.2.5). Because of the assumptions in model (5.2.4)

$$V_f \sim \sigma^2 \chi^2_{m-1} = \frac{1}{\tau}\, \chi^2_{m-1}$$

where χ^2_{m-1} is a chi-squared variable based on $(m-1)$ degrees of freedom. So its

density is given by

$$p(V_f|\tau,\xi,z) = (\frac{1}{2})^{\frac{1}{2}\upsilon_2}\ (V_f)^{\frac{1}{2}\upsilon_2-1}\ \frac{1}{\Gamma(\upsilon_2/2)}\ \tau^{\frac{1}{2}\upsilon_2}\exp\left[-\frac{\tau V_f}{2}\right]$$

where the degrees of freedom, υ_2 , is $m-1$. Using the result of Theorem 5.1,

$$p(\bar{Y}' \mid \tau,\xi,z,V_f) = \frac{1}{\sqrt{2\pi}}\ \frac{\tau^{\frac{1}{2}}}{\sqrt{C_\xi}}\ \exp\left[\frac{-(\bar{Y}'-\mu')^2\tau}{2C_\xi}\right]$$

where $\mu' = \beta^T\hat{\gamma}$ and

$$C_\xi = (\beta^T S_1^{-1}\beta + \frac{1}{m}). \tag{5.2.8}$$

Substituting for $p(\bar{Y}' \mid \tau,\xi,z,V_f)$ and $p(V_f \mid \tau,\xi,z)$ in expression (5.2.7) gives the following expression for $L(\xi)$

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}}\int_0^\infty \tau^{\frac{1}{2}(\upsilon_2+1)}\ e^{-\frac{1}{2}E_\xi\tau}\pi(\tau \mid \xi,z)\ d\tau \tag{5.2.9}$$

where $E_\xi = \dfrac{(\bar{Y}'-\mu')^2}{C_\xi} + V_f$.

5.2.1 A prior density for τ

The residual sum of squares *RSS* of the calibration experiment is given by

$$RSS(\alpha) = \sum_{i=1}^{N}(Y_i - \hat{f}(x_i))^2$$

where $\hat{f}$ is the spline smoother for the calibration data set and α is the smoothing parameter. Wahba (1978) suggested a formula for estimating the variance σ^2

$$\hat{\sigma}^2 = \frac{\sum(Y_i-\hat{f}(x_i))^2}{n-trA(\alpha)} = \frac{RSS(\alpha)}{n-trA(\alpha)}$$

where there are n distinct knots $x_1, x_2, \ldots, x_n$ and $A(\alpha)$ is the *hat matrix.* The term $n - trA(\alpha)$ is called the *equivalent degrees of freedom.* Now *RSS* is a sufficient statistic for τ so the posterior density of τ is given by

$$\pi(\tau \mid \xi,z) \propto \pi(\tau)p(RSS \mid \tau) \tag{5.2.10}$$

using Lemma 1.4.1., Box and Tiao (1973). Here $\pi(\tau)$ denotes the prior density of τ. One could adopt a non-informative prior for τ which corresponds to taking

$$\pi(\tau) \propto \frac{1}{\tau}$$

but as there was a reasonable amount of prior information on τ it was decided to adopt an informative prior for τ, namely a Gamma distribution with parameters $\frac{1}{2}g$

and $\frac{1}{2}h$, i.e. $Ga(\frac{1}{2}g,\frac{1}{2}h)$. The prior density for τ is given by

$$\pi(\tau) = \frac{(\frac{1}{2}h)^{\frac{1}{2}g}\,\tau^{g/2-1}\exp(-\frac{1}{2}h\tau)}{\Gamma(\frac{1}{2}g)} \quad g > 0,\ h > 0,\ \tau > 0. \qquad (5.2.11)$$

The vague prior referred to above corresponds to letting $g \to 0$ and $h \to 0$. This Gamma density function has a unique maximum at $(g-2)/h$, $E(\tau) = g/h$, and $\mathrm{Var}(\tau) = 2g/h^2$. In the case of $h = 1$ and g a positive integer the density (5.2.11) is that of a χ^2 distribution with g degrees of freedom so g was taken to be the *equivalent degrees of freedom.*

$$g = n - trA(\alpha). \qquad (5.2.12)$$

Since the density given by equation (5.2.11) has a maximum at $(g-2)/h$ and the simulated data sets were generated by using a particular value for $\tau(\tau_s)$ the value of h was taken to be

$$h = (g-2)/\tau_s = \frac{n-trA(\alpha)-2}{\tau_s}. \qquad (5.2.13)$$

In the case of τ being unknown, a good estimate of τ from past data could be used in equation (5.2.13) instead of τ_s. This was done with the antibiotic assay data set (see Section 5.6).

If one adopts a prior density for τ given by equation (5.2.11) then using expression (5.2.10) the posterior density of τ is given by

$$\pi(\tau \mid \xi,z) \propto \tau^{g/2-1}\, e^{(-\frac{1}{2}h\tau)}\, p(RSS \mid \tau)$$

If one now substitutes for $\pi(\tau \mid \xi,z)$ in the expression for $L(\xi)$ in (5.2.9), this gives

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \int_0^\infty \tau^{\frac{1}{2}(v_2+g-1)}\, e^{-\frac{1}{2}\tau(E_t+h)} p(RSS \mid \tau)\, d\tau \qquad (5.2.14)$$

5.2.2 Calculation of C_ξ

The term C_ξ is given by equation (5.2.8) and it is not difficult to compute because of the banded nature of S_1. The matrix S_1 is defined as

$$S_1 = \alpha\Omega + B^{\mathrm{T}}WB$$

where B and Ω are $n \times n$ matrices defined by

$$B_{ij} = \beta_j(x_i) \quad \text{and} \quad \Omega_{ij} = \int_{-\infty}^{\infty} \beta_i''(x)\beta_j''(x)dx$$

Here $\beta_j(x_i)$ is the value of the jth B-spline at the knot x_i (see Chapter 3, Section 3.1.3). The matrices Ω and B are banded matrices and W is diagonal, so the $n \times n$

matrix S_1 is a symmetric band matrix of bandwidth four. Let S_1 have the Cholesky decomposition $S_1 = LL^T$ where L is a lower-triangular band matrix of bandwidth four.

Let $\beta^T = (\beta_1(\xi)\ \beta_2(\xi) \ldots\ldots \beta_n(\xi))$ then C_ξ can be written as

$$\beta^T(LL^T)^{-1}\beta + \frac{1}{m}$$

$$= (L^{-1}\beta)^T(L^{-1}\beta) + \frac{1}{m} \qquad \text{since } (L^T)^{-1} = (L^{-1})^T.$$

Suppose we let $L^{-1}\beta = \mathbf{z}$, then $L\mathbf{z} = \beta$. This band-limited lower triangular linear system can be solved easily and quickly given L and β. Then C_ξ is given by $\mathbf{z}^T\mathbf{z} + 1/m$. The elements of the column vector β can be calculated using the equations (3.1.10) and (3.1.11). For example, consider the calculation of $\beta_i(\xi)$ $i=3,\ldots,(n-2)$

$$\beta_i(\xi) = N_{i-2,4}(\xi).$$

In order to generate $N_{i-2,4}(\xi)$ using the recursive formula (3.1.10), one requires to firstly generate $N_{i-2,1}(\xi)$, $N_{i-1,1}(\xi)$, $N_{i,1}(\xi)$, $N_{i+1,1}(\xi)$; then $N_{i-2,2}(\xi)$, $N_{i-1,2}(\xi)$, $N_{i,2}(\xi)$ and finally $N_{i-2,3}(\xi)$ and $N_{i-1,3}(\xi)$. The first four N's are either 0 or 1 since

$$N_{j,1}(\xi) = \begin{cases} 1 & \text{if } \xi \in [x_j, x_{j+1}) \\ 0 & \text{otherwise} \end{cases}$$

If $\xi \in (x_j, x_{j+1})$ $j=2,3,\ldots,(n-2)$, there are non-zero contributions to β from only four B-splines, namely $\beta_{j-1}(\xi)$, $\beta_j(\xi)$, $\beta_{j+1}(\xi)$ and $\beta_{j+2}(\xi)$. If $\xi = x_j$ $j=2,\ldots,(n-2)$ this reduces to three B-splines since $\beta_{j+2}(\xi) = 0$ in this case. Also if $\xi \in [x_1, x_2)$ or $\xi \in [x_{n-1}, x_n]$ there are non-zero contributions from at most three B-splines. Hence for any given ξ, $x_1 \le \xi \le x_n$, the column vector β has at most four non-zero entries.

Section 5.3 considers the application of Imhof's results to obtaining an approximate expression for $\pi(\tau \mid \xi, \mathbf{z})$ which results in an expression for $L(\xi)$ which is a double semi-infinite integral. The applications of Eagleson-Buckley and Solomon-Stephens approximations to the distribution of $\varepsilon^T D \varepsilon$ result in an expression for $L(\xi)$ which is a single semi-infinite integral. (See Sections 5.4 and 5.5 respectively). Having found an expression for $L(\xi)$ it is a relatively easy task to obtain the posterior density $\pi(\xi \mid \mathbf{Y}', \mathbf{z})$ since from expression (5.2.6)

$$\pi(\xi \mid \mathbf{Y}', \mathbf{z}) \propto \pi(\xi \mid \mathbf{x})L(\xi)$$
$$\pi(\xi)L(\xi)$$

depending on whether the calibration experiment is random or controlled.

5.3 IMHOF'S METHOD

Imhof's results (1961) were derived for quadratic forms of the form $(\mathbf{x}+\mu)^T F(\mathbf{x}+\mu)$ which involve non-central χ^2-variables. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ be a column vector which has a multivariate normal distribution with mean vector $\mathbf{0}$ and variance matrix $\sum$. Let $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$ be a constant vector. If $\sum$ is non-singular one can express Q in the form

$$Q = \sum_{r=1}^{m} \lambda_r \, \chi^2_{h_r;\delta_r^2}.$$

The λ_r are the distinct non-zero eigenvalues of $F\sum$, the h_r their respective orders of multiplicity, the δ_r certain linear combinations of $\mu_1, \ldots, \mu_n$ and the $\chi^2_{h_r;\delta_r^2}$ are independent χ^2-variables with h_r degrees of freedom and non-centrality parameter δ_r^2. If $\varphi(t)$ is the characteristic function of Q then $\varphi(t)$ is given by

$$\varphi(t) = \prod_{1}^{m} (1-2i\lambda_r t)^{-\frac{1}{2}h_r} \exp\left[i \sum_{1}^{m} \frac{\delta_r^2 \lambda_r t}{1-2i\lambda_r t} \right]. \qquad (5.3.1)$$

The probability density function $g(x)$ associated with Q is given by

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi(t) dt.$$

However, Imhof explains that to obtain $g(x)$ by integration of the inversion formula is hopeless except in the case of $m = 1$. Imhof therefore puts forward a formula for obtaining the cumulative distribution function $F(x)$ of the variable Q based on a formula derived by Gil-Pelaez (1951), namely

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} t^{-1} \operatorname{Im}\{e^{-itx} \varphi(t)\} \, dt$$

where $\operatorname{Im}(z)$ denotes the imaginary part of z and $\varphi(t)$ is the characteristic function of Q given by equation (5.3.1). The result is the following;

$$P(Q > x) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\sin\theta(u)}{u\rho(u)} \, du \qquad \text{where}$$

$$\theta(u) = \frac{1}{2} \sum_{1}^{m} [h_r \tan^{-1}(\lambda_r u) + \delta_r^2 \lambda_r u (1+\lambda_r^2 u^2)^{-1}] - \frac{1}{2} xu$$

$$\rho(u) = \prod_{1}^{m} (1+\lambda_r^2 u^2)^{\frac{1}{4}h_r} \exp\{\tfrac{1}{2} \sum_{1}^{m} (\delta_r \lambda_r u)^2 / (1+\lambda_r^2 u^2)\}$$

The probability density $g(x)$ of the quadratic form Q is given by

$$g(x) = \frac{1}{\pi} \int_0^{\infty} \frac{\cos\theta(u)}{\rho(u)} \, du. \qquad (5.3.2)$$

As stated in Section 5.1, we are concerned with the quadratic form $Q_n = \varepsilon^T D \varepsilon$ which involves central χ^2 random variables so $\delta_r^2 = 0$ in the expressions for $\theta(u)$ and $\rho(u)$. In this case

$$\theta(u) = \frac{1}{2}\sum_1^m h_r \tan^{-1}(\lambda_r u) - \frac{1}{2}xu \quad \text{and} \quad \rho(u) = \prod_1^m (1+\lambda_r^2 u^2)^{\frac{1}{4}h_r}$$

Let $\varepsilon^* = \varepsilon/\sigma$ then $\varepsilon^* \sim \mathcal{N}_n(0, I_n)$. Let $Q_n^* = \varepsilon^{*T} D \varepsilon^* = \varepsilon^T D \varepsilon/\sigma^2$. The density function $h(x)$ associated with Q_n^* is given by

$$h(x) = \frac{1}{\pi}\int_0^\infty \frac{\cos\left[\frac{1}{2}\sum_1^m h_r \tan^{-1}(\lambda_r u) - \frac{1}{2}xu\right]}{\prod_1^m (1+\lambda_r^2 u^2)^{\frac{1}{4}h_r}}\, du \qquad (5.3.3)$$

where λ_r are the distinct non-zero eigenvalues of matrix D. Let us consider these eigenvalues in more detail.

The positive definite $n \times n$ matrix D is defined by $D = (I-A)^2$ where A is the *hat matrix*. Let the eigenvalues of matrix A be represented by $\lambda_{1A}, \lambda_{2A} \ldots\ldots \lambda_{nA}$ and the eigenvalues of matrix D be represented by $\lambda_{1D}, \lambda_{2D} \ldots\ldots \lambda_{nD}$. Silverman (1984) showed that the eigenvalues of A are given approximately by

$$\lambda_{1A} = \lambda_{2A} = 1$$

$$\lambda_{iA} = [1 + c_0\alpha(i-1.5)^4]^{-1} \qquad i=3,4\ldots\ldots n$$

where $c_0 = \frac{\pi^4}{n}\left[\int_a^b \hat{f}(t)^{\frac{1}{4}} dt\right]^{-4}$. Here $\hat{f}(t)$ is an estimate of the local density of the points x_i calculated using the fast algorithm of Silverman (1982) on a range $[a,b]$ where $a = x_1 - \frac{1}{2}n^{-1}(x_n - x_1)$ and $b = x_n + \frac{1}{2}n^{-1}(x_n - x_1)$. Hence the eigenvalues of D are given approximately by

$$\lambda_{iD} = 0 \quad i=1,2$$

$$\lambda_{iD} = [1 - (1 + c_0\alpha(i-1.5)^4)^{-1}]^2 \qquad i=3,4\ldots\ldots n \qquad (5.3.4)$$

Returning to the equation for $h(x)$ given by (5.3.3) $h(x)$ is now given by

$$h(x) = \frac{1}{\pi}\int_0^\infty \frac{\cos\left[\frac{1}{2}\sum_3^n \tan^{-1}(\lambda_{iD} u) - \frac{1}{2}xu\right]}{\prod_3^n (1+\lambda_{iD}^2 u^2)^{\frac{1}{4}}}\, du \qquad (5.3.5)$$

where the λ_{iD} are defined by equation (5.3.4). The above equation (5.3.5) defines the probability density function for $Q_n^* = \varepsilon^{*T} D \varepsilon^* = (\varepsilon^T D \varepsilon)\tau$ and as stated in section 5.1 of this chapter, the distribution of *RSS* can be approximated by the

distribution of $\varepsilon^T D\varepsilon$, provided the bias term $\mathbf{f}^T D\mathbf{f}$ is not too large. Hence the probability density function associated with *RSS* for a given τ is given by

$$p(RSS \mid \tau) = \frac{\tau}{\pi}\int_0^\infty \frac{\cos\left[\frac{1}{2}\sum_3^n \tan^{-1}(\lambda_{iD}u) - \frac{1}{2}(RSS)\tau u\right]}{\rho(u)}\,du$$

where $\rho(u) = \prod_3^n (1+\lambda_{iD}^2 u^2)^{\frac{1}{4}}$. Substituting for $p(RSS \mid \tau)$ in equation (5.2.14) gives the following expression for $L(\xi)$

$$\frac{1}{\pi\sqrt{C_\xi}}\int_0^\infty\int_0^\infty \tau^a\, e^{-\frac{1}{2}(E_\xi^*\tau)}\,\frac{\cos\left[\frac{1}{2}\sum_3^n \tan^{-1}(\lambda_{iD}u) - \frac{1}{2}(RSS)\tau u\right]}{\rho(u)}\,du\,d\tau$$

where $E_\xi^* = \dfrac{(\bar{Y}'-\mu')^2}{C_\xi} + V_f + h$ and $a = \frac{1}{2}(\nu_2 + g + 1)$. Making a substitution $y = E_\xi^*\tau/2$ gives

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}}\,\frac{1}{E_\xi^{*a+1}}\int_0^\infty\int_0^\infty y^a e^{-y}\,\frac{\cos\theta(u,y)}{\rho(u)}\,du\,dy$$

where $\theta(u,y) = \frac{1}{2}\sum_{i=3}^n \tan^{-1}(\lambda_{iD}u) - RSS\,uy/E_\xi^*$. The eigenvalues λ_{iD} $(i=3,\ldots,n)$ were estimated using the BATHSPLINE program (Silverman and Watters, 1984).

It was decided to evaluate the double integral by applying a different one-dimensional method to each dimension. In particular

$$\int_0^\infty\int_0^\infty f(u,y)\,du\,dy \approx \sum_{j=1}^{N^*} w_j\left[\int_0^\infty f(u,y_j)\,du\right]$$

$$\text{where}\qquad f(u,y_j) = \frac{1}{\sqrt{C_\xi}}\left[\frac{1}{E_\xi^*}\right]^{a+1} y_j^a\,\frac{\cos\theta(u,y_j)}{\rho(u)}$$

With the Gauss-Laguerre quadrature based on N^* abscissae y_j and *normal* weights, the summation $\sum_{j=1}^{N^*} w_j\, g(y_j)$ approximates the integral

$$\int_{a'}^\infty e^{-by}\, g(y)\, dy \qquad (b>0)$$

and is exact for any function of the form $g(y) = \sum_{j=0}^{2N^*-1} c_j y^j$. Here $b=1$, $a'=0$, and $g(y_j) = \int_0^\infty f(u, y_j)\, du$. It will be recalled that $\rho(u)$ is defined as

$$\rho(u) = \prod_{i=3}^n (1+\lambda_{iD}^2 u^2)^{\frac{1}{4}}.$$

The function, $\rho(u)$, increases monotonically towards infinity so the numerical integration of $f(u,y_j)$ with respect to u can be carried out over a finite range $0 \le u \le U^*$. The degree of approximation obtained in evaluating $\int_0^\infty f(u, y_j)\, du$ numerically will depend (apart from rounding-off errors) on two sources of error

(i) the *error of integration* resulting from the use of an approximate rule for computing $\int_0^{U^*} \frac{\cos\theta(u,y)}{\rho(u)}\, du$

(ii) the *error of truncation* $t_u = \int_{U^*}^\infty \frac{\cos\theta(u,y)}{\rho(u)}\, du$

Although it is not feasible to obtain an upper bound for the error of integration, it is possible to obtain an upper bound for the truncation error, t_u. For a given y,

$$|t_u| \le \int_{U^*}^\infty \frac{1}{\prod_{i=3}^{n} |\lambda_{iD}|^{\frac{1}{2}}\, u^{m/2}}\, du$$

where m is the number of non-zero eigenvalues λ_{iD}. So

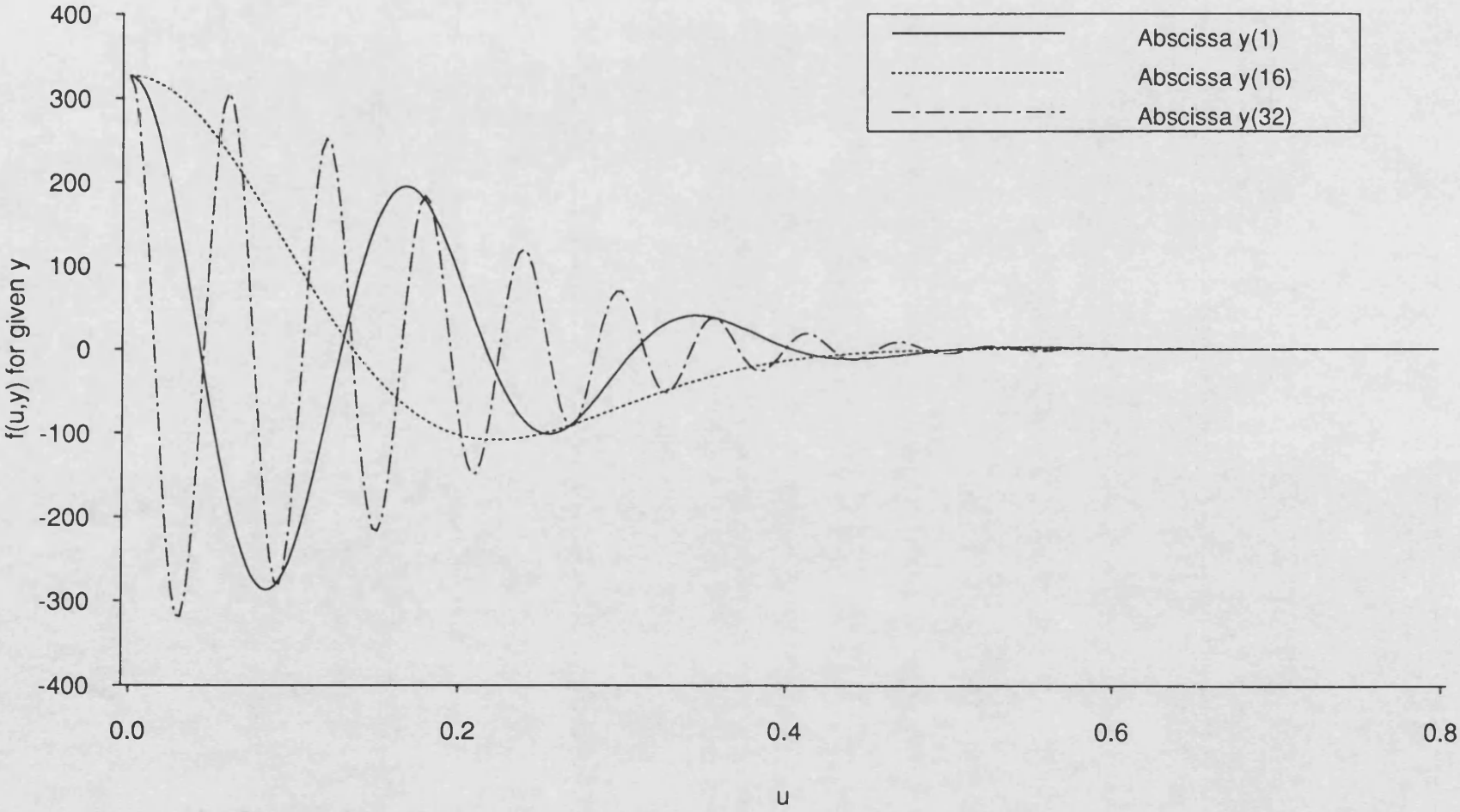
$$|t_u| \le \frac{1}{\prod_{i=3}^{n} |\lambda_{iD}|^{\frac{1}{2}}} \frac{1}{k(U^*)^k}$$

where $k = m/2 - 1$. In the analysis of the data sets in this chapter the upper bound for the truncation error was taken to be 10^{-12} and the value of U^* calculated. The values for the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets are 2.2714, 2.2750, 2.6258, 2.2847 and 2.0777 respectively. Fig 5.1 shows $[c\cos\theta(u,y)]/\rho(u)$ plotted against u for the abscissae $y_1 = 0.044489$, $y_{16} = 19.855861$ and $y_{32} = 111.751398$. The underlying data set are the Weibull data. Here c is a scaling constant of magnitude $e^{15} \times 10^{-4}$. Because of the oscillatory nature of $\frac{\cos\theta(u,y)}{\rho(u)}$ the integral

$$\int_0^{U^*} \frac{\cos\theta(u,y)}{\rho(u)}\, du$$

was evaluated using the Nag routine D01AKF which is an adaptive integrator, especially suited to oscillating non-singular integrands. It uses the Gauss 30-point and Kronrod 61-point rules and is based upon the QUADPACK routine QAG (Piessens, De Doncker-Kapenga, Uberhuber, and Kahaner, 1983).

Fig.5.1: Imhof's integrand for various abscissae



The interval $[0,U^*]$ is repeatedly divided into a number of subintervals and the integration rule is applied separately to each subinterval. The idea of an adaptive routine is that the distribution of abscissae (u values in this case) is adapted to the shape of the integrand with more subdivisions in the neighbourhood of peaks.

The Nag routine D01BBF/D01BAX was used to return the *normal* weights w_j and abscissae y_j for a Gauss-Laguerre quadrature formula with $N^* = 32$ so an approximation to the double integral

$$\frac{1}{\sqrt{C_\xi}} \left[\frac{1}{E_\xi^*}\right]^{a+1} \int_0^\infty \int_0^\infty y^a e^{-y} \frac{\cos\theta(u,y)}{\rho(u)}\, du\, dy$$

was calculated and hence an evaluation of $L(\xi)$ was made.

Substituting for $L(\xi)$ in equation (5.2.6) gives

$$\pi(\xi \mid \mathbf{Y}',\mathbf{z}) = \frac{C_I \pi(\xi)}{\sqrt{C_\xi}} \left[\frac{1}{E_\xi^*}\right]^{a+1} \int_0^\infty \int_0^\infty y^a e^{-y} \frac{\cos\theta(u,y)}{\rho(u)}\, du\, dy \quad (5.3.6)$$

where C_I is a normalising constant.

In Section 5.7, Imhof's method is applied to the five data sets Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic using a uniform prior for ξ. For this prior, the posterior density $\pi(\xi \mid \mathbf{Y}',\mathbf{z})$ given by equation (5.3.6) was evaluated at 201 values of ξ over the calibration range $[x_1, x_n]$ and the constant C_I was evaluated using a Nag routine D01GAF based on the method of Gill and Miller (1972). The Nag routine uses third-order finite differences and integrates a function which is specified numerically at four or more points.

5.4 THE EAGLESON-BUCKLEY APPROXIMATION

Let us consider the quadratic form $Q_n = \varepsilon^{\mathrm{T}} D \varepsilon$ where $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 I_n)$. From equation (5.1.5) the mean and variance of Q_n are $\sigma^2 \mathrm{tr}D$ and $2\sigma^4 \mathrm{tr}D^2$ respectively. Eagleson and Buckley (1988) considered the *normalised* quadratic form Q^* defined by

$$\frac{Q_n - E(Q_n)}{\sqrt{\mathrm{Var}(Q_n)}} = \frac{\varepsilon^{\mathrm{T}} D \varepsilon - \sigma^2 \mathrm{tr}D}{\sqrt{2\sigma^4 \mathrm{tr}D^2}} = \frac{\varepsilon^* D \varepsilon^* - \mathrm{tr}D}{\sqrt{2\,\mathrm{tr}D^2}}$$

where $\varepsilon^* = \varepsilon/\sigma$. With the Eagleson-Buckley approximation, the distribution function of Q^* is approximated by a function $G(u_0\,; x)$ which is close to the two-term Edgeworth expansion:

$$\Phi(x) + \frac{1}{6} u_0 (1-x^2) Z(x). \quad (5.4.1)$$

Here $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$ and $Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. It is possible to obtain

useful approximate representations of a distribution given by $g(x)$ in terms of the normal density function $Z(x)$. If in the expansion, terms of equal order in $n^{-\frac{1}{2}}$ are collected together and arranged in ascending order the expansion is called an *Edgeworth expansion* for g. The leading terms are

$$g(x) = [1 + \frac{1}{6}\sqrt{\beta_1}H_3(x) + \frac{1}{24}(\beta_2 - 3)H_4(x) + \frac{1}{72}\beta_1 H_6(x) + ...]\, Z(x)$$

where $H_j(x)$ are the Hermite polynomials. The two-term Edgeworth expansion is given by

$$g(x) = [1 - \frac{1}{6}\sqrt{\beta_1}(x^3 - 3x)]Z(x)$$

and

$$\int_{-\infty}^{x} g(t)dt = \Phi(x) + \frac{1}{6}\sqrt{\beta_1}(1-x^2)Z(x). \qquad (5.4.2)$$

Here $\sqrt{\beta_1}$ is the coefficient of skewness for the distribution defined by $g(x)$ and is defined as

$$\sqrt{\beta_1} = \frac{E(X - E(X))^3}{[\mathrm{Var}(X)]^{3/2}}.$$

Here the random variable of interest is Q^* and $\sqrt{\beta_1}$ is given by $E[(Q^*)^3]$ since $E(Q^*) = 0$ and $\mathrm{Var}(Q^*) = 1$. Letting $Q_n = \varepsilon^{\mathrm{T}} D \varepsilon$ gives

$$\sqrt{\beta_1} = E(Q^{*3}) = \frac{E(Q_n - E(Q_n))^3}{[\mathrm{Var}(Q_n)]^{3/2}}. \qquad (5.4.3)$$

Using the results given in equation (5.1.5) gives

$$\sqrt{\beta_1} = \frac{8\sum_k \lambda_k^3}{[2\sum \lambda_k^2]^{3/2}} = \frac{2\sqrt{2}\sum_k \lambda_k^3}{(\sum_k \lambda_k^2)^{3/2}}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of matrix $\sigma^2 D$. Comparing equation (5.4.2) with expression (5.4.1) reveals that

$$u_0 = \sqrt{\beta_1} = \frac{2\sqrt{2}\sum_k \lambda_k^3}{(\sum_k \lambda_k^2)^{3/2}}. \qquad (5.4.4)$$

So the distribution of Q^* is approximated by a function $G(u_0; x)$ where u_0 is given by equation (5.4.4) above.

Eagleson and Buckley then go on to consider a similar expansion for the χ_ν^2 distribution. They approximate the distribution of the normalised χ_ν^2 random

variable (χ_ν^*) given by

$$\chi_\nu^* = \frac{\chi_\nu^2 - \nu}{\sqrt{2\nu}}$$

by the function $G(2\sqrt{2}\nu^{-\frac{1}{2}}; x)$. By choosing $u_0 = 2\sqrt{2}\nu^{-\frac{1}{2}}$ the distribution of Q^* can be approximated by the distribution of χ_ν^*. This implies, using equation (5.4.4) that the degrees of freedom ν are given by

$$\nu = (\sum_k \lambda_k^2)^3 / (\sum_k \lambda_k^3)^2 = \frac{(trD^2)^3}{(trD^3)^2} \tag{5.4.5}$$

Hence the distribution of Q^* is approximated by $c\chi_\nu^2 + d$ where

$$c = \frac{1}{\sqrt{2\nu}} \quad \text{and} \quad d = -\sqrt{\frac{\nu}{2}} \tag{5.4.6}$$

and ν is defined in equation (5.4.5). The approximation of Q^* by $c\chi_\nu^2 + d$ is equivalent to equating the first three moments of Q^* and $c\chi_\nu^2 + d$ since

$$E(c\chi_\nu^2 + d) = E(\chi_\nu^*) = E(Q^{*}) = 0$$

$$\text{Var}(c\chi_\nu^2 + d) = \text{Var}(\chi_\nu^*) = \text{Var}(Q^*) = 1$$

Using equations (5.4.5) and (5.4.6) the third central moment of $c\chi_\nu^2 + d$ or χ_ν^* is given by

$$\begin{aligned} E(\chi_\nu^* - E(\chi_\nu^*))^3 &= c^3 E[(\chi_\nu^2 - \nu)^3] \\ &= 8c^3\nu = 2\sqrt{2}\nu^{-\frac{1}{2}} \\ &= 2\sqrt{2} \sum_k \lambda_k^3 / (\sum \lambda_k^2)^{3/2} = \sqrt{\beta_1} \end{aligned}$$

which is the third central moment of Q^* (see equation (5.4.3))

As explained in Section 5.1 of this chapter, provided the bias term $\mathbf{f}^T D \mathbf{f}$ is not too large the distribution of RSS can be approximated by the distribution of $\boldsymbol{\varepsilon}^T D \boldsymbol{\varepsilon}$. Hence the distribution of RSS/σ^2 can be approximated by the distribution of $\boldsymbol{\varepsilon}^{*T} D \boldsymbol{\varepsilon}^*$ so that approximately

$$\frac{\frac{RSS}{\sigma^2} - \text{tr}D}{\sqrt{2\,\text{tr}D^2}} \sim \frac{1}{\sqrt{2\nu}} \chi_\nu^2 - \sqrt{\frac{\nu}{2}}$$

where ν is given by equation (5.4.5). Let $\tau = 1/\sigma^2$ then approximately

$$RSS\,\tau \sim c^* \chi_\nu^2 + d^*$$

$$\text{where} \quad c^* = \frac{\text{tr}D^3}{(trD^2)} \quad \text{and} \quad d^* = \left[\text{tr}D - \frac{(trD^2)^2}{\text{tr}D^3}\right]. \tag{5.4.7}$$

Suppose the random variable X has a χ^2-distribution based on ν degrees of freedom. Then the p.d.f. of X is

$$g(x) = \frac{1}{2^{\frac{1}{2}\nu}} \frac{1}{\Gamma(\frac{1}{2}\nu)} x^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}x} \qquad x > 0$$

$$= 0 \qquad \text{otherwise.}$$

Let $Y = RSS\,\tau$ then the density function of Y is given by

$$p(Y) \propto \left[\frac{Y-d^*}{c^*}\right]^{\frac{1}{2}\nu-1} \exp\left[-\frac{(Y-d^*)}{2c^*}\right] \qquad Y > d^*$$

$$0 \qquad\qquad\qquad \text{otherwise}$$

So the density function of RSS for a given τ is given by

$$p(RSS|\tau) \propto \tau \left[\frac{RSS\,\tau-d^*}{c^*}\right]^{\frac{1}{2}\nu-1} \exp\left[-\frac{(RSS\,\tau-d^*)}{2c^*}\right] \qquad RSS > d^*/\tau$$

$$0 \qquad\qquad \text{otherwise}$$

Substituting for $p(RSS|\tau)$ in equation (5.2.14) gives the following result for the predictive density function $L(\xi)$.

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \int_{d'}^{\infty} \tau^a (\tau-d')^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}E_\xi'\tau}\, d\tau$$

where $E_\xi' = \dfrac{(\bar{Y}'-\mu')^2}{C_\xi} + V_f + \dfrac{1}{c'} + h$, $a = \frac{1}{2}(\nu_2+g+1)$, $d' = d^*/RSS$ and $c' = c^*/RSS$. Making a substitution $x = E_\xi'\tau/2$ gives

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \int_{x'}^{\infty} \left[\frac{x}{E_\xi'}\right]^a \left[\frac{1}{E_\xi'}\right] \left[\frac{2x}{E_\xi'}-d'\right]^{\frac{1}{2}\nu-1} e^{-x}\, dx$$

$$= \frac{1}{\sqrt{C_\xi}} \left[\frac{1}{E_\xi'}\right]^p \int_{x'}^{\infty} x^a (x-x')^{\frac{1}{2}\nu-1} e^{-x}\, dx \qquad (5.4.8)$$

where $x' = E_\xi' d'/2$ and $p = a+\frac{1}{2}\nu$. Tables of integrals (Prudnikov, Brychkov and Marichev, 1986) give

$$\int_{x'}^{\infty} x^a (x-x')^{\frac{1}{2}\nu-1} e^{-x}\, dx$$

$$= \Gamma(\tfrac{1}{2}\nu)\,(x')^p\, e^{-x'}\, \Psi(\tfrac{1}{2}\nu, \tfrac{1}{2}(\nu_2+\nu+g+3); x')$$

where $\Psi(b,c;z) = \dfrac{\Gamma(1-c)}{\Gamma(1+b-c)} M(b,c;z) + \dfrac{\Gamma(c-1)}{\Gamma(b)} z^{1-c} M(1+b-c, 2-c; z)$

The functions $M(b,c;z)$ and $M(1+b-c,2-c;z)$ are called *confluent hypergeometric functions* (Abramowitz and Stegun, 1965). The function $M(b,c;z)$ is defined as follows:

$$M(b,c;z) = 1 + \frac{b}{c}z + \frac{b(b+1)}{c(c+1)}\frac{z^2}{2!} + \frac{b(b+1)(b+2)}{c(c+1)(c+2)}\frac{z^3}{3!} + \cdots\cdots\cdots$$

It is a convergent series for all b, c and z if $c \neq -n$ and $b \neq -m$ where n and m are positive integers. If $c \neq -n$ and $b = -m$ then it is a polynomial of degree m in z.

For the integral given in expression (5.4.8) $b = \frac{1}{2}\nu$ and $c = \frac{1}{2}(\nu_2+\nu+g+3)$ and $b, c > 0$. Also $(1+b-c) = -\frac{1}{2}(\nu_2+g+1)$ and $(2-c) = \frac{1}{2}(1-\nu_2-\nu-g)$. The degrees of freedom, ν, are those associated with RSS and are given by equation (5.4.5). It will not in general be an integer. Also g is defined in equation (5.2.12) and it will not in general be an integer so b, c, $2-c$ and $1+b-c$ will not have integer values. It therefore follows that $M(\frac{1}{2}\nu,\frac{1}{2}(\nu_2+\nu+g+3);x')$ and $M(-\frac{1}{2}(\nu_2+g+1),\frac{1}{2}(1-\nu_2-\nu-g);x')$ will be convergent series. It follows therefore that the predictive density function $L(\xi)$ is a well-defined function of ξ and is also integrable for all values of ξ in $[x_1, x_n]$.

Because of the numerical difficulties of evaluating $\Gamma(x)$ for negative x, it was decided to evaluate $L(\xi)$ using numerical integration techniques since

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \left[\frac{1}{E'_\xi}\right]^p \int_{x'}^{\infty} f(x)\, dx \qquad (5.4.9)$$

where $f(x) = x^{\frac{1}{2}(\nu_2+g+1)}\,(x-x')^{\frac{1}{2}\nu-1}\,e^{-x}$. Because of the size of ν and g in the data sets $f(x)$ was evaluated as $\exp\left[\frac{1}{2}(\nu_2+g+1)\ln x + (\frac{1}{2}\nu-1)\ln(x-x') - x\right]$. The Nag routine D01BAF was used for evaluating $\int_{x'}^{\infty} f(x)\,dx$, in particular the Gauss-Laguerre quadrature formula with 64 abscissae. The Gauss-Laguerre formula is exact for any function of the form

$$f(x) = e^{-bx} \sum_{i=0}^{2N^*-1} c_i x^i$$

where N^* is the number of points of the Gaussian quadrature. Here b=1 and N^* was taken to be 64 for all the data sets.

Substituting for $L(\xi)$ in equation (5.2.6) gives

$$\pi(\xi \mid \mathbf{Y}', \mathbf{z}) = C_I \frac{\pi(\xi)}{\sqrt{C_\xi}} \left[\frac{1}{E_\xi'}\right]^p \int_{x'}^{\infty} x^a\,(x-x')^{\frac{1}{2}\nu-1}\,e^{-x}\,dx \qquad (5.4.10)$$

where $a = \frac{1}{2}(\nu_2 + g + 1)$, $p = a + \frac{1}{2}\nu$ and C_I is a normalising constant.

In Section 5.7, the approximation described in this section is applied to the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets using a variety of prior densities for ξ. For each chosen prior, the posterior density $\pi(\xi \mid \mathbf{Y}', \mathbf{z})$ defined in equation (5.4.10) was evaluated at 201 values of ξ over the calibration range $[x_1, x_n]$ and the normalising constant C_I evaluated using the Nag routine D01GAF.

5.5 THE SOLOMON-STEPHENS APPROXIMATION

With the Solomon-Stephens approximation, the distribution of the quadratic form $Q_n = \varepsilon^T D \varepsilon$ is approximated by $c(\chi_p^2)^d$. The parameters c, p and d are found by equating the first three moments of the two distributions. Unlike the Eagleson-Buckley approximation, there is no theoretical justification for the Solomon-Stephens approximation. From Johnson and Kotz (1970), the moments about the origin of a χ_p^2 random variable are given by

$$\mu_r'(\chi_p^2) = E[(\chi_p^2)^r] = 2^r \frac{\Gamma(\frac{1}{2}p+r)}{\Gamma(\frac{1}{2}p)} \qquad r = 1,2,3,\ldots .$$

Let $p/2 = v^*$ and suppose $W = c(\chi_p^2)^d$ then

$$\mu_r'(\chi_p^2) = 2^r \frac{\Gamma(v^*+r)}{\Gamma(v^*)}$$

$$E(W) = E[c(\chi_p^2)^d] = c\mu_d' = c2^d \frac{\Gamma(v^*+d)}{\Gamma(v^*)}$$

$$E(W^2) = E[c^2(\chi_p^2)^{2d}] = c^2 \mu_{2d}' = c^2 \frac{2^{2d}\Gamma(v^*+2d)}{\Gamma(v^*)}$$

$$E(W^3) = E[c^3(\chi_p^2)^{3d}] = c^3 \mu_{3d}' = c^3 2^{3d} \frac{\Gamma(v^*+3d)}{\Gamma(v^*)}.$$

Let $f_1(v^*, d) = E(W^2)/[E(W)]^2$ and $f_2(v^*, d) = E(W^3)/[E(W)]^3$ then

$$f_1(v^*, d) = \frac{\Gamma(v^*)\Gamma(v^*+2d)}{[\Gamma(v^*+d)]^2} \tag{5.5.1}$$

$$f_2(v^*, d) = \frac{[\Gamma(v^*)]^2 \Gamma(v^*+3d)}{[\Gamma(v^*+d)]^3}. \tag{5.5.2}$$

Johnson and Kotz (1970) prove that the cumulants of $Q_n = \varepsilon^T D \varepsilon$ are given by

$$\kappa_s = \sigma^{2s} 2^{s-1}(s-1)! \ \mathrm{tr}(D^s).$$

The first, second and third moments about the origin of the distribution of Q_n are given by

$$\mu_1' = E(Q_n) = \kappa_1 = \sigma^2 \operatorname{tr} D$$

$$\mu_2' = E(Q_n^2) = \kappa_2 + \kappa_1^2 = 2\sigma^4 \operatorname{tr} D^2 + \sigma^4 (\operatorname{tr} D)^2$$

$$\mu_3' = E(Q_n^3) = \kappa_3 + 3\kappa_2 \kappa_1 + \kappa_1^3$$

$$= 8\sigma^6 \operatorname{tr} D^3 + 6\sigma^6 \operatorname{tr} D^2 \operatorname{tr} D + \sigma^6 [\operatorname{tr} D]^3.$$

Hence $$\frac{E(Q_n^2)}{[E(Q_n)]^2} = 1 + \frac{2 \operatorname{tr} D^2}{(\operatorname{tr} D)^2} \tag{5.5.3}$$

$$\frac{E(Q_n^3)}{[E(Q_n)]^3} = \frac{8 \operatorname{tr} D^3 + 6 \operatorname{tr} D^2 \operatorname{tr} D + (\operatorname{tr} D)^3}{(\operatorname{tr} D)^3}. \tag{5.5.4}$$

Equating (5.5.1) with (5.5.3) and (5.5.2) with (5.5.4) gives

$$\frac{\Gamma(\nu^*)\Gamma(\nu^*+2d)}{[\Gamma(\nu^*+d)]^2} = 1 + \frac{2 \operatorname{tr} D^2}{(\operatorname{tr} D)^2} \tag{5.5.5}$$

$$\frac{[\Gamma(\nu^*)]^2 \Gamma(\nu^*+3d)}{[\Gamma(\nu^*+d)]^3} = 1 + \frac{6 \operatorname{tr} D^2}{(tr D)^2} + \frac{8 \operatorname{tr} D^3}{(tr D)^3}. \tag{5.5.6}$$

Suppose the right-hand side of equation (5.5.3) is denoted by c_1 and the right-hand side of equation (5.5.4) is denoted by c_2. The equations (5.5.5) and (5.5.6) can be written as

$$f_1(\nu^*, d) = c_1$$

$$f_2(\nu^*, d) = c_2.$$

These were solved by an iterative method involving the minimisation of the expression

$$F = (f_1 - c_1)^2 + (f_2 - c_2)^2$$

subject to $\frac{1}{2} \leq \nu^* \leq \frac{1}{2}(N-1)$ and $d \geq 0$ where N is the number of observations in the calibration data set z. The algorithm used was a quasi-Newton algorithm by Gill and Murray (1976, Nag routine E04JAF) where the minimisation is subject to bounds on the variables , which in this case are ν^* and d. At the beginning of this section, $2\nu^*$ was taken to be equal to p which is the degrees of freedom of the approximating χ^2 variable. For all the data sets considered, the value of $2\nu^*$ which minimised F was the *equivalent degrees of freedom* defined by equation (5.2.12).

Having found the values of d and ν^* which satisfy equations (5.5.5) and (5.5.6) the parameter c can be found by using the fact that

$$E(Q_n) = E(W) = \frac{c\, 2^d \Gamma(\nu^*+d)}{\Gamma(\nu^*)} = \sigma^2 \operatorname{tr} D$$

therefore $$c = \frac{\sigma^2 \mathrm{tr} D\,\Gamma(\nu^*)}{2^d \Gamma(\nu^*+d)}. \tag{5.5.7}$$

As mentioned in previous sections of this chapter, if the bias term $\mathbf{f}^T D\mathbf{f}$ is not large then the distribution of RSS can be approximated by the distribution of $Q_n = \boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$.

So since $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon} \sim c(\chi_p^2)^d$ approximately where c is given by equation (5.5.7), then approximately $RSS \sim c(\chi_p^2)^d$. Let $c^* = c/\sigma^2$ then

$$\frac{\mathrm{RSS}}{\sigma^2} \sim c^*(\chi_p^2)^d \tag{5.5.8}$$

where $$c^* = \frac{\mathrm{tr} D\,\Gamma(\nu^*)}{2^d\,\Gamma(\nu^*+d)}. \tag{5.5.9}$$

Let $\tau = 1/\sigma^2$ then from expression (5.5.8)

$$\mathrm{RSS}\,\tau \sim c^*(\chi_p^2)^d.$$

Suppose the random variable X has an χ^2-distribution based on ν degrees of freedom. Then the p.d.f. of X is

$$g(x) = \frac{1}{2^{\frac{1}{2}\nu}} \frac{1}{\Gamma(\frac{1}{2}\nu)} x^{\frac{1}{2}\nu-1} \exp(-\tfrac{1}{2}x) \quad x > 0$$

$$= 0 \qquad \text{otherwise.}$$

Let $Y = RSS\,\tau$ then the density function of Y is given by

$$P(Y) \propto (\frac{Y}{c^*})^{\frac{p}{2d}-1} \exp(-\tfrac{1}{2}[\frac{Y}{c^*}]^{1/d}).$$

So the density function of RSS for a given τ is given by

$$P(RSS|\tau) \propto \tau \left[\frac{RSS\,\tau}{c^*}\right]^{\frac{p}{2d}-1} \exp\left[-\tfrac{1}{2}\left[\frac{RSS\,\tau}{c^*}\right]^{1/d}\right].$$

Substituting for $p(RSS \mid \tau)$ in equation (5.2.14) gives the following expression for the predictive density function $L(\xi)$

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \int_0^\infty \tau^a \exp(\frac{-K_\xi \tau}{2}) \exp(-\tfrac{1}{2}[\frac{\tau}{c'}]^{1/d})\, d\tau$$

where $a = \frac{1}{2}(g+\nu_2+\frac{p}{d}-1)$, $K_\xi = \frac{(\bar{Y}'-\mu')^2}{C_\xi} + V_f + h$ and $c' = c^*/RSS$. Making the substitution $x = K_\xi \tau/2$ gives

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \left[\frac{1}{K_\xi}\right]^{a+1} \int_0^\infty x^a \exp[-x-K_\xi^* x^{1/d}]\, dx \tag{5.5.10}$$

where $K_\xi^* = \frac{1}{2}[(\frac{2}{K_\xi c'})^{\frac{1}{d}}]$. So the predictive density function $L(\xi)$ is defined by

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}} \left[\frac{1}{K_\xi}\right]^{a+1} \int_0^\infty f(x)dx$$

where $f(x) = x^a \exp[-x-K_\xi^* x^{1/d}]$. Because of the size of p and g for the data sets, $f(x)$ was evaluated as $\exp[\, a\ln x - x - K_\xi^* x^{1/d}]$. The semi-infinite integral above was not evaluated using a Gauss-Laguerre quadrature formula as was the case for the Eagleson-Buckley approximation. This was not a suitable formula to use because of the presence of the term $\exp[-K_\xi^* x^{1/d}]$. There was a build-up of round-off error making the results totally unreliable. Instead an automatic adaptive procedure was used similar to that used in Section 5.3, namely D01AMF. As with the routine D01AKF used in Section 5.3, this Nag routine is based on QUADPACK (Piessens, De Doncker, Uberhuber and Kahaner 1983) in particular the routine DQAGI. The routine D01AMF calculates an approximation to $\int_0^\infty f(x)\, dx$ by first transforming $[0,\infty)$ to $(0,1)$ by using the identity

$$\int_0^\infty f(x)dx = \int_0^1 f\left[\frac{1-t}{t}\right] \frac{1}{t^2} dt.$$

Then an adaptive procedure based on the Gauss 7-point and Kronrod 15-point rules is used on the transformed integral. The algorithm is described in De Doncker (1978).

Substituting for $L(\xi)$ in equation (5.2.6) using equation (5.5.10), the posterior density $\pi(\xi \,|\, \mathbf{Y}', \mathbf{z})$ is given by

$$\pi(\xi \mid \mathbf{Y}', \mathbf{z}) = \frac{C_I \pi(\xi)}{\sqrt{C_\xi}} \left[\frac{1}{K_\xi}\right]^{a+1} \int_0^\infty x^a \exp[-x-K_\xi^* x^{\frac{1}{d}}]\, dx \quad (5.5.11)$$

where C_I is a normalising constant. In Section 5.7, the approximation described in this section is applied to the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets using a variety of prior densities for ξ. For each chosen prior, the posterior density $\pi(\xi \mid \mathbf{Y}', \mathbf{z})$ defined in equation (5.5.11) was evaluated at 201 values of ξ over the calibration range $[x_1, x_n]$ and the normalising constant C_I evaluated using the Nag routine D01GAF.

5.6 THE ANTIBIOTIC ASSAY DATA

The model considered appropriate for this data was first considered in Chapter 3 (equation 3.3.3). The model is given by

$$D_{ij} = \mu_i + \varepsilon_{ij} \qquad i=1,2,\ldots 7 \qquad j=1,2,3\ldots..n_i$$

$$= f(c_i) + \varepsilon_{ij} \tag{5.6.1}$$

where D_{ij} is the jth clearance diameter at level i of C, and f is a natural cubic *interpolating* spline with knots $\{c_i\}$. It is assumed that $\varepsilon_{ij} \sim N(0,\sigma^2)$. The residual sum of squares associated with the calibration experiment is given by

$$RSS = \sum_{i=1}^{7} \sum_{j=1}^{n_i} (D_{ij} - \hat{f}(c_i))^2$$

where $\hat{f}$ is the cubic *interpolating* spline which interpolates the points $(c_i, \bar{D}_{i\cdot})$ $i=1,2,\ldots 7$. Here $\bar{D}_{i\cdot} = \sum_{j=1}^{n_i} D_{ij}/n_i = \hat{f}(c_i)$ so the residual sum of squares is given by

$$RSS = \sum_i \sum_j (D_{ij} - \bar{D}_{i\cdot})^2.$$

Because of the normality and independence assumptions about the ε_{ij} it follows that

$$\sum_{i=1}^{7} \sum_{j=1}^{n_i} \frac{(D_{ij} - \bar{D}_{i\cdot})^2}{\sigma^2} \sim \chi_k^2$$

$$RSS/\sigma^2 \sim \chi_k^2$$

where $k = \sum_i (n_i - 1)$. The model associated with the prediction stage of the calibration experiment is given by

$$D_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,\ldots m \tag{5.6.2}$$

where f is a natural cubic interpolating spline. It is not necessary with this non-parametric model to approximate the distribution of RSS since

$$p(RSS|\tau) \propto \tau^{\frac{1}{2}k} e^{-\frac{1}{2}RSS\tau} \tag{5.6.3}$$

and the following theorem will prove that the predictive density function $L(\xi)$ is a non-central Student density.

Theorem 5.2

Suppose that *a priori*, ξ is independent of (γ,σ^2), the prior distribution of γ is locally uniform, i.e. $p(\gamma \mid \sigma^2) \propto c$ where c is a constant and suppose that the prior distribution of $\tau = 1/\sigma^2$ is $Ga(\frac{1}{2}g,\frac{1}{2}h)$ where g and h are defined by equations (5.2.12) and (5.2.13) respectively. Let β_i represent the ith B-spline and let $\beta^T = (\beta_1(\xi)\ \beta_2(\xi) \ldots\ldots \beta_n(\xi))$. Then the predictive density function $L(\xi)$ is given by

$$L(\xi) = St(\nu,\ \beta^{\mathrm{T}}\hat{\gamma},\ \frac{V}{\nu}C_{\xi}) \tag{5.6.4}$$

where $\nu = (m-1) + k + g$, $C_{\xi} = (\beta^{\mathrm{T}}S_1^{-1}\beta + \frac{1}{m})$

$S_1 = B^{\mathrm{T}}WB$ and $\hat{\gamma} = S_1^{-1}B^{\mathrm{T}}W\overline{\mathbf{D}}$

$V = RSS + \sum_{j=1}^{m}(D_j' - \overline{D}')^2 + h$

The vector $\overline{\mathbf{D}}$ is defined as $\overline{\mathbf{D}}^{\mathrm{T}} = (\overline{D}_1.\ \overline{D}_2.\ \overline{D}_3.\ \ldots\ldots\ \overline{D}_7.)$ and $\overline{D}' = \sum_{j=1}^{m} D_j'/m$.

Proof. Consider firstly the following model for the calibration experiment

$$Y_i = f(x_i) + \varepsilon_i$$

where f is a natural cubic smoothing spline with distinct knots $\{x_i\}$, $\mathrm{Var}(Y_i) = \sigma^2/w_i$ and it is assumed that the ε_i are i.i.d. normal random variables with mean 0. Assume that the model at the prediction stage is given by

$$Y_j' = f(\xi) + \varepsilon_j'$$

where again f is a natural cubic smoothing spline and ε_j' are i.i.d. normal random variables with mean 0 and variance σ^2. Applying Corollary 5.1 gives

$$\{\overline{Y}'|\sigma^2,\xi,\mathbf{z},V_f\} \sim N(\beta^{\mathrm{T}}\hat{\gamma},\ \beta^{\mathrm{T}}S^{-1}\beta + \frac{\sigma^2}{m}) \tag{5.6.5}$$

where $\hat{\gamma} = \frac{1}{\sigma^2}S^{-1}B^{\mathrm{T}}W\mathbf{Y}$, $S = (B^{\mathrm{T}}WB + \alpha\Omega)/\sigma^2$ and $\mathbf{Y}^{\mathrm{T}} = (\overline{Y}_1\,\overline{Y}_2\,\overline{Y}_3 \ldots \overline{Y}_n)$.
Now if one takes the smoothing parameter α to be zero f is a natural cubic *interpolating* spline with knots $\{x_i\}$ and in this case

$$\{\overline{Y}'|\sigma^2,\xi,\mathbf{z},V_f\} \sim N(\beta^{\mathrm{T}}\hat{\gamma},\beta^{\mathrm{T}}S^{-1}\beta + \frac{\sigma^2}{m})$$

where $\hat{\gamma}$ is as given above but S is given by $S = B^{\mathrm{T}}WB/\sigma^2$. Let $S_1 = \sigma^2 S$ then result (5.6.5) becomes

$$\{\overline{Y}'|\sigma^2,\xi,\mathbf{z},V_f\} \sim N(\beta^{\mathrm{T}}\hat{\gamma},\ \sigma^2(\beta^{\mathrm{T}}S_1^{-1}\beta + 1/m)).$$

Applying this to the models (5.6.1) and (5.6.2) where f is a natural cubic interpolating spline gives

$$\{\overline{D}'|\sigma^2,\xi,\mathbf{z},V_f\} \sim N(\beta^{\mathrm{T}}\hat{\gamma},\ C_{\xi}\sigma^2) \tag{5.6.6}$$

where $C_{\xi} = \beta^{\mathrm{T}}S_1^{-1}\beta + \frac{1}{m}$, $S_1 = B^{\mathrm{T}}WB$ and $\hat{\gamma} = S_1^{-1}B^{\mathrm{T}}W\overline{\mathbf{D}}$

Let $\tau = 1/\sigma^2$. By Bayes theorem

$$\pi(\tau|\xi,V_f,\mathbf{z}) \propto \pi(\tau|\xi,\mathbf{z})\,p(V_f|\tau,\xi,\mathbf{z})$$
$$= \pi(\tau|\xi,\mathbf{z})\,p(V_f|\tau) \qquad (5.6.7)$$

where $V_f = \sum_{j=1}^{m}(D_j'-\bar{D}')^2$. As a function of τ

$$p(V_f|\tau) \propto \tau^{\frac{1}{2}(m-1)}e^{-\frac{1}{2}V_f\tau}. \qquad (5.6.8)$$

From equation (5.6.3) $p(RSS|\tau) \propto \tau^{\frac{1}{2}k}e^{-\frac{1}{2}RSS\tau}$ where $k = \sum_i(n_i-1)$. The prior density of τ is assumed to be $Ga(\frac{1}{2}g,\frac{1}{2}h)$. So the posterior density of τ, $\pi(\tau|\xi,\mathbf{z})$ is given by

$$\pi(\tau|\xi,\mathbf{z}) \propto \tau^{\frac{1}{2}k+\frac{1}{2}g-1}\,e^{-\frac{1}{2}(RSS+h)\tau}.$$

Substituting in equation (5.6.7) for $p(V_f|\tau)$ and $\pi(\tau|\xi,\mathbf{z})$ gives

$$\pi(\tau|\xi,V_f,\mathbf{z}) \propto \tau^{\frac{1}{2}\nu-1}\,e^{-\frac{1}{2}V\tau} \qquad (5.6.9)$$

where $\nu = (m-1) + k + g$ and $V = RSS + V_f + h$. The predictive density function $L(\xi)$ is given by

$$L(\xi) = p(\bar{D}'|\xi,\mathbf{z},V_f) = \int_0^\infty p(\bar{D}'|\tau,\xi,\mathbf{z},V_f)\pi(\tau|\xi,\mathbf{z},V_f)\,d\tau.$$

Using results (5.6.6) and (5.6.9) gives

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}}\int_0^\infty \tau^{\frac{1}{2}(\nu-1)}e^{-K_\xi\tau/2}\,d\tau$$

where $K_\xi = (\bar{D}'-\beta^{\mathrm{T}}\hat{\gamma})^2/C_\xi + V$. Making a substitution $K_\xi\tau/2 = u$, the integral simplifies to give

$$L(\xi) \propto \frac{1}{\sqrt{C_\xi}}\left[\frac{1}{K_\xi}\right]^{\frac{1}{2}(\nu+1)}\int_0^\infty u^{\frac{1}{2}(\nu-1)}\,e^{-u}\,du$$
$$\propto \frac{1}{\sqrt{C_\xi}}\left[\frac{1}{K_\xi}\right]^{\frac{1}{2}(\nu+1)}$$
$$\propto \frac{1}{(VC_\xi)^{\frac{1}{2}}\left[1+\frac{(\bar{D}'-\beta^{\mathrm{T}}\hat{\gamma})^2}{VC_\xi}\right]^{\frac{1}{2}(\nu+1)}}.$$

This is the density function (non-normalised) of a non-central Student t-distribution based on ν degrees of freedom, centred at $\beta^{\mathrm{T}}\hat{\gamma}$ and with scale factor $\frac{VC_\xi}{\nu}$, i.e.

$$L(\xi) = St(\nu,\ \beta^{\mathrm{T}}\hat{\gamma},\ \frac{VC_\xi}{\nu}).$$

The posterior density of ξ is given by

$$\pi(\xi|\mathbf{D}',\mathbf{z}) \propto \pi(\xi)L(\xi)$$

since this is a controlled calibration experiment. It is therefore a relatively easy matter to calculate the posterior densities $\pi(\xi|\mathbf{D}',\mathbf{z})$ for a given prior density for ξ, $\pi(\xi)$, since $L(\xi)$ is a non-central Student t density.

5.6.1 Application to the antibiotic assay data

A uniform prior was assumed for ξ, in particular

$$\pi(\xi) = \frac{1}{(c_7-c_1)} \qquad c_1 \leq \xi \leq c_7$$

$$= 0 \qquad \text{otherwise.}$$

The posterior density for ξ, $\pi(\xi|\mathbf{D}',\mathbf{z})$ is given by

$$\pi(\xi|\mathbf{D}',\mathbf{z}) = \frac{C_I}{(VC_\xi)^{\frac{1}{2}}\left[1+\frac{(\bar{D}'-\beta^{\mathrm{T}}\hat{\gamma})^2}{VC_\xi}\right]^{\frac{\nu+1}{2}}} \tag{5.6.10}$$

where C_I is a normalising constant. Equation (5.2.13) gives the value of h as $(g-2)/\tau_s = (g-2)\sigma_s^2$ where g is the degrees of freedom associated with the residual sum of squares of the calibration experiment and σ_s^2 is the known value of σ^2. Whereas the value σ_s is known for the simulated data sets, it is not known here but can be estimated from past data. There were two other independent data sets derived from similar calibration experiments with the *same* log concentration levels $c_1, c_2, c_3 \ldots\ldots, c_7$. Estimates of σ^2 from the two calibration data sets are given by

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{7}\sum_{j=1}^{n_{1i}}(D_{1ij}-\bar{D}_{1i\cdot})^2}{\sum_{i=1}^{7}(n_{1i}-1)} = \frac{\sum_{i=1}^{7}\sum_{j=1}^{n_{1i}}(D_{1ij}-\bar{D}_{1i\cdot})^2}{k_1}$$

$$\hat{\sigma}_2^2 = \frac{\sum_{i=1}^{7}\sum_{j=1}^{n_{2i}}(D_{2ij}-\bar{D}_{2i\cdot})^2}{\sum_{i=1}^{7}(n_{2i}-1)} = \frac{\sum_{i=1}^{7}\sum_{j=1}^{n_{2i}}(D_{2ij}-\bar{D}_{2i\cdot})^2}{k_2}$$

where $\sum_{i=1}^{7}\sum_{j=1}^{n_i}(D_{ij}-\bar{D}_{i\cdot})^2$ is the pooled *within* sums of squares for the calibration

data and k is the pooled degrees of freedom. The differences in the two estimates of σ^2 were non-significant, so a pooled estimate, of σ^2, $\hat{\sigma}_p^2$, was formed using

$$\hat{\sigma}_p^2 = \frac{k_1 \hat{\sigma}_1^2 + k_2 \hat{\sigma}_2^2}{(k_1 + k_2)} . \qquad (5.6.11)$$

The value of g is therefore taken to be $k_1 + k_2$ and so the value of h is $\hat{\sigma}_p^2(k_1 + k_2 - 2)$. As there was only *one* measurement of clearance diameter D' made at the prediction stage, m= 1 in model (5.6.2) so $\sum_{j=1}^{m}(D_j' - \bar{D}')^2 = 0$. Hence v is $k_1 + k_2 + 0 + g = 2(k_1 + k_2)$ and V is given by

$$RSS + \sum_{j=1}^{m}(D_j' - \bar{D}')^2 + h = RSS + h = RSS + \hat{\sigma}_p^2(k_1 + k_2 - 2)$$

where $\hat{\sigma}_p^2$ is given by equation (5.6.11).

The posterior density $\pi(\xi|D', \mathbf{z})$ was evaluated at 301 values of ξ over the calibration range $[c_1, c_7]$ and the normalising constant, C_I , evaluated using the Nag routine D01GAF. Fig 5.2 shows the posterior density curves for D' = 12.0, D' = 17.0 and D' = 23.0. The density curves for D' = 12.0 and D' = 23.0 are highly skewed which is to be expected given the nature of the calibration data and since all the relevant information pertaining to the value of ξ is contained within the calibration range $[c_1, c_7]$. An obvious point estimate for ξ in these two cases is the posterior mode. However, the posterior median was also calculated for the distributions $\pi(\xi \mid D'$=12.00, $\mathbf{z})$ and $\pi(\xi \mid D'$=23.00, $\mathbf{z})$.

Sixty estimates of ξ were calculated, both point and interval estimates. Table 5.1 gives the mean absolute errors using both the posterior median and the posterior mode.

Table 5.1 also gives details of the interval estimates obtained for ξ from the posterior distributions of ξ. Both symmetrical and highest posterior density interval estimates were obtained for ξ using a 90% probability level and these are denoted in Table 5.1 as SY and HP respectively. Interval estimates of ξ were termed *successful* if the true value ξ lay within the interval. Study of the results shows that the posterior mode is marginally better as a point estimator of ξ. Also the highest posterior density interval estimates are better than the symmetrical interval estimates particularly with reference to the number of *successful* estimates. This arises because of the high degree of skewness in the posterior distributions of ξ where the true log concentration of antibiotic is close to or equal to c_1 or c_7. In fact five out of the seven *unsuccessful* symmetrical interval

Fig.5.2: Posterior densities for antibiotic assay data

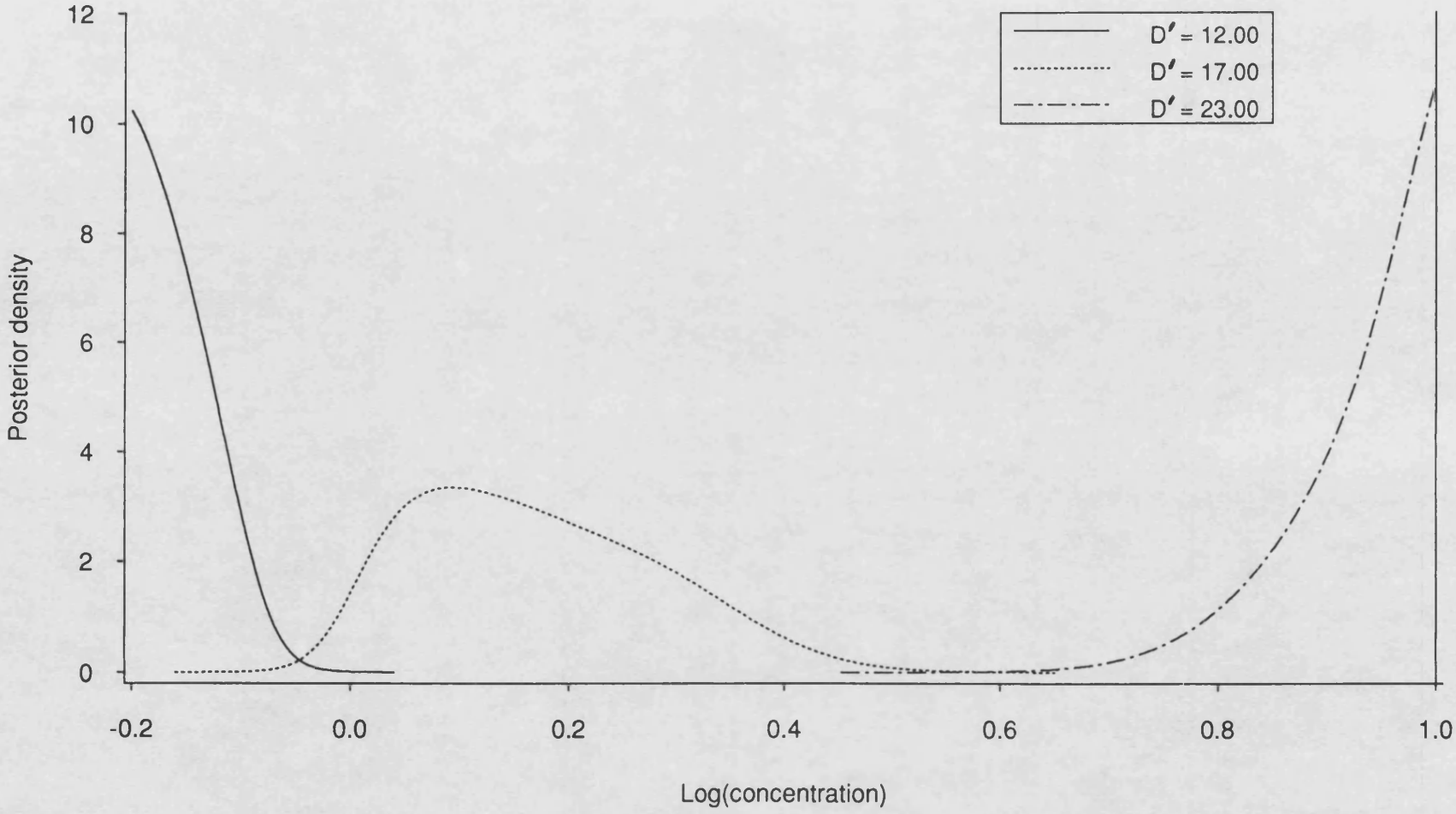


Table 5.1 Point estimates and interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Mean absolute error	Mode	0.061
	Median	0.067
Maximum width of interval estimates	SY	0.46
	HP	0.46
Minimum width of interval estimates	SY	0.13
	HP	0.11
Mean width of interval estimates	SY	0.303
	HP	0.290
No. of successful estimates	SY	53
	HP	57
Total no. of estimates	60	
Calibration range	−0.22 to 1.00	

estimates occurred where the true log concentration of antibiotic was $c_7 = 1.00$. The upper limits of four of these five unsuccessful estimates were 0.99. A comparison of these results with the corresponding results of Chapters 3 and 4 will be made in Chapter 7.

5.7 APPLICATION TO THE NON-LINEAR SIMULATED DATA SETS

Imhof's method, Eagleson-Buckley's (E-B) approximation and Solomon-Stephens' (S-S) approximation detailed in Sections 5.3, 5.4 and 5.5 respectively were applied to the Weibull, Gompertz, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets, the same data sets as those considered in Chapters 3 and 4. (See Figs. 3.7 - 3.11). The teeth data are not analysed here because an earlier analysis in Chapter 3 indicated that a linear model was a more suitable model for these data sets. These data sets will be analysed in Chapter 7 using the predictive density approaches of Hoadley (1970) and Aitchison and Dunsmore (1975).

Let us consider first the application of the E-B and S-S approximations to the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets.

5.7.1 Comparison of the Eagleson-Buckley and Solomon-Stephens approximations

Suppose the model of the prediction stage is given by

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3.$$

To assess the approximations, twenty values of X were chosen to cover the calibration range of each data set, in fact the same twenty values as those chosen to assess the methods of Chapters 3 and 4.

With the E-B approximation the distribution of RSS/σ^2, where RSS is the residual sum of squares for the calibration experiment, is approximated by $c^* \chi_\nu^2 + d^*$ where c^* , d^* and ν are given respectively by equations (5.4.7) and (5.4.5). Table 5.2 gives the values of c^*,d^* and ν for each of the five data sets. Using the Solomon-Stephens approximation involves approximating the distribution of RSS/σ^2 by $c^*(\chi_p^2)^d$ where c^* is given by equation (5.5.9) and p and d are solutions of equations (5.5.5) and (5.5.6). The values of c^*, p and d are given in Table 5.2 also. It is interesting to note that the E-B approximation is based on a much smaller number of degrees of freedom than the S-S approximation. For the E-B approximation the posterior density of ξ, $\pi(\xi|\, \mathbf{Y}', \mathbf{z})$ is given by equation (5.4.10) whereas for the S-S approximation it is given by equation (5.5.11).

Table 5.2 Details of the approximations to the distribution of RSS/σ^2

Data set	E-B approximation			S-S approximation		
	c^*	d^*	v	c^*	d	p
Gompertz	0.990	0.397	74.16	0.0064	1.726	224.54
Weibull	0.990	0.399	74.13	0.0064	1.726	224.48
Preece-Baines	0.985	0.557	72.16	0.0065	1.725	219.36
Bleasdale-Nelder	0.990	0.404	74.07	0.0064	1.726	224.32
Asymptotic	0.993	0.287	75.53	0.0064	1.726	228.08

Three different prior densities for ξ were assumed:

(i) a uniform density on $[x_1, x_n]$ so that

$$\pi(\xi) = \frac{1}{(x_n - x_1)} \qquad x_1 \leq \xi \leq x_n$$

$$= 0 \qquad \text{otherwise.}$$

(ii) a triangular density on $[x_1, x_n]$ so that

$$\pi(\xi) = \frac{4(\xi - x_1)}{(x_n - x_1)^2} \qquad x_1 \leq \xi \leq \tfrac{1}{2}(x_1 + x_n)$$

$$= \frac{4(x_n - \xi)}{(x_n - x_1)^2} \qquad \tfrac{1}{2}(x_1 + x_n) \leq \xi \leq x_n$$

$$= 0 \qquad \text{otherwise.}$$

(iii) a truncated non-central Student density so that

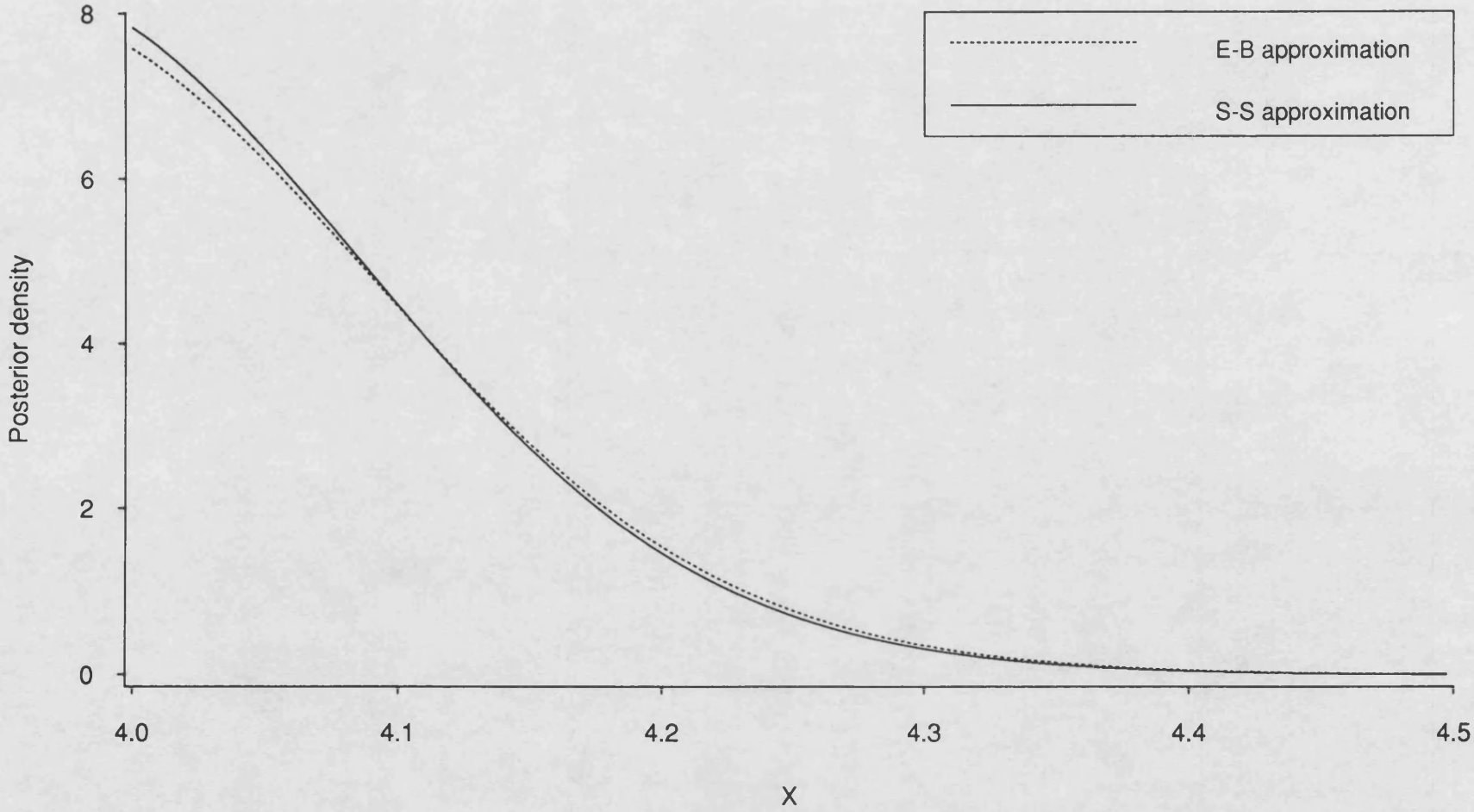
$$\pi(\xi) = \frac{C_s}{[(1+\frac{1}{N})S_{xx}]^{\frac{1}{2}} \left[1 + \frac{(\xi - \bar{x})^2}{(1+\frac{1}{N})S_{xx}}\right]^{N/2}} \qquad x_1 \leq \xi \leq x_n$$

$$= 0 \qquad \text{otherwise.}$$

Here N is the number of observations (x_i, Y_i) constituting the calibration data, $\bar{x} = \frac{\Sigma x_i}{N}$, $S_{xx} = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{(N-1)}$ and C_S is a normalising constant.

The uniform density for ξ was used with all five data sets whereas the triangular density was used with the Gompertz, Preece-Baines and Bleasdale-Nelder data sets only. The truncated Student prior was used with the Weibull and Asymptotic data sets only. Fig. 5.3 shows the posterior distributions of ξ for the Preece-Baines data set assuming a uniform prior for ξ and using both approximations. The true value, ξ, equals x_1 which is 4.00. The figure clearly shows that there are only small differences in the two posterior distributions and both are very skewed to the right. One would expect the posterior distributions to be skewed in this way since all the information pertaining to the value of ξ is contained within the calibration range $[4.00, 17.80]$.

Fig.5.3: Comparison of posterior densities



Point estimates for ξ were calculated using both the posterior mode and posterior median. Tables 5.3 and 5.4 give the mean absolute error for each of the five data sets when using a uniform prior density. Examination of these two tables shows that both approximations give good estimates for ξ even when the number of replications, m, is only one. This latter statement is further confirmed by Fig. 5.4 which shows the median, $\hat{\xi}_{\text{median}}$ and the mode, $\hat{\xi}_{\text{mode}}$, plotted against the true value, ξ, for the Weibull data with $m = 1$ and using the S-S approximation. The 45° line represents zero error. Tables 5.3 and 5.4 indicate that there is virtually no difference in the two approximations, their mean absolute errors agreeing to one decimal place. The tables also indicate that the use of the posterior mode rather than the posterior median produces a lower mean absolute error for all the data sets except the Gompertz data set and the Weibull data set with $m = 1$. The difference is most marked for the Preece-Baines data set. Fig. 5.5 shows the modal estimate ($\hat{\xi}_{\text{mode}}$) and the median estimate ($\hat{\xi}_{\text{median}}$), plotted against the true value, ξ, for the Asymptotic data set using the E-B approximation. The 45° line represents zero error. It shows clearly a good feature of this predictive density approach using an approximation to the distribution of RSS, namely that where the spline is rather flat and the calibration data are therefore not very informative concerning the value of ξ, the errors are larger . For the Asymptotic data set this occurs when $\xi > 11$ approximately (see Fig. 3.11).

By way of illustration, Tables 5.5 and 5.6 give the mean absolute error when using the triangular prior for ξ and Tables 5.7 and 5.8 give the mean absolute error when using the Student prior for ξ for a selected number of data sets. Again these tables demonstrate that as far as point estimates are concerned, the two approximations are equally good. Let $\hat{\xi}_u$ be the posterior median when one assumes a uniform prior density for ξ and $\hat{\xi}_{Tr}$ be the posterior median when one assumes a triangular prior density for ξ. Fig. 5.6 shows $\hat{\xi}_u$ and $\hat{\xi}_{Tr}$ plotted against the true value, ξ, for the Gompertz data set using the E-B approximation. The prior mean is $\frac{1}{2}(x_1+x_2)$ which equals 6.715. Again the 45° line represents zero error. Fig. 5.6 shows that $\hat{\xi}_{Tr}$ can be thought of as a shift of $\hat{\xi}_u$ towards the prior mean, the shift being largest when the calibration data are least informative. For this data set this is when $0 \leq \xi \leq 3$ and $\xi > 11$ approximately (see Fig. 3.7).

With regard to interval estimates for ξ, 90% symmetrical interval estimates for ξ were constructed from the posterior distributions of ξ. An interval estimate was termed *successful* if the true value ξ lay inside the interval. Also 90% highest posterior density (HPD) interval estimates for ξ were constructed from the posterior distributions of ξ.

Table 5.3 Mean absolute error when using the mode and a uniform prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Gompertz	0.38	0.38	0 - 13.43
Weibull			
m=1	1.72	1.72	
m=2	0.86	0.86	10 - 79.52
m=3	0.70	0.69	
Preece-Baines	0.12	0.12	4 - 19.8
Bleasdale-Nelder	3.20	3.20	20 - 178
Asymptotic	1.11	1.11	1 - 21.74

Table 5.4 Mean absolute error when using the median and a uniform prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Gompertz	0.35	0.35	0 - 13.43
Weibull			
m=1	1.62	1.62	
m=2	0.86	0.86	10 - 79.52
m=3	0.74	0.73	
Preece-Baines	0.17	0.17	4 - 19.8
Bleasdale-Nelder	3.32	3.31	20 - 178
Asymptotic	1.15	1.14	1 - 21.74

Fig.5.4: Comparison of point estimates
(Weibull data set, S-S approximation)

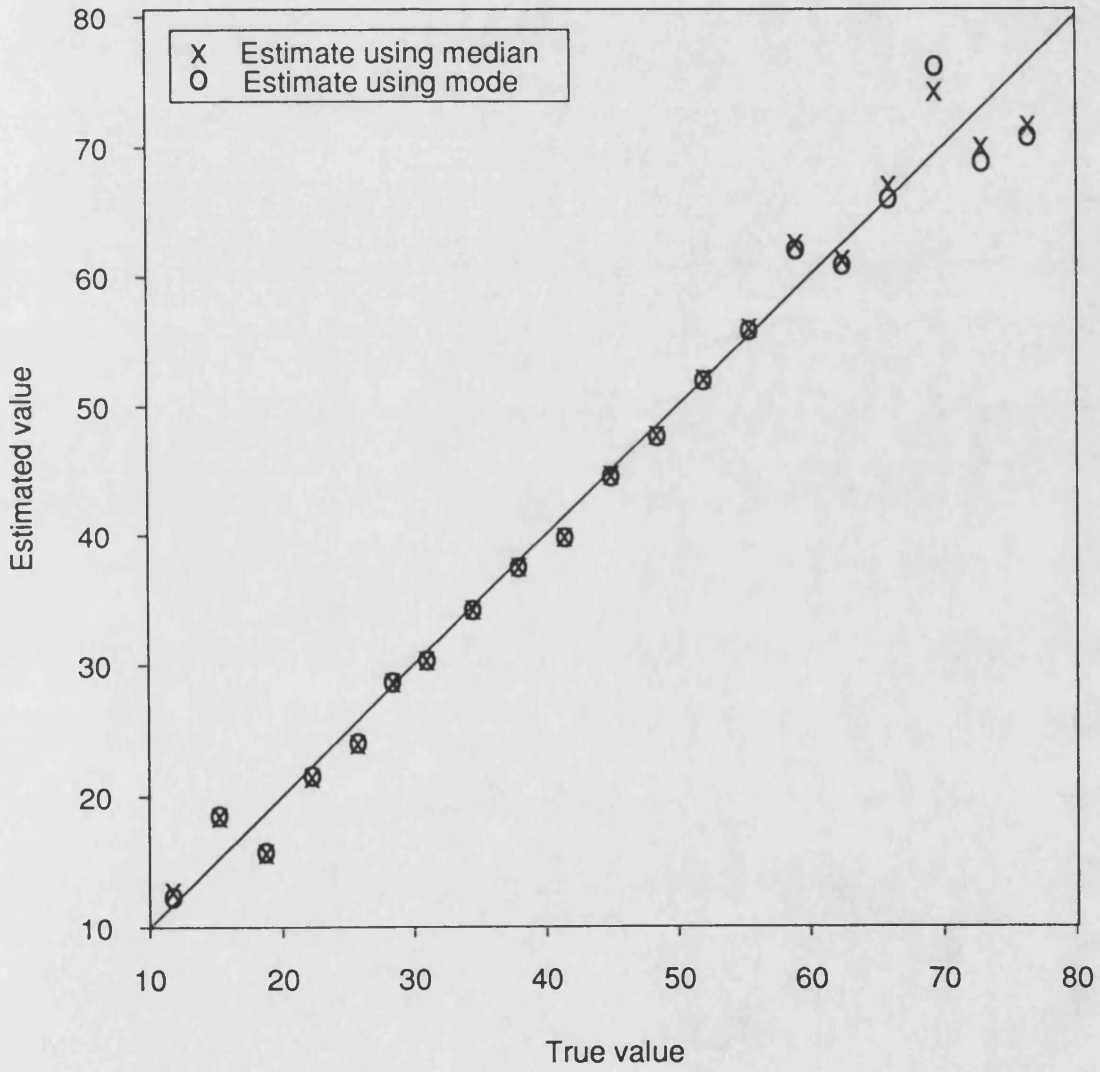


Fig.5.5: Comparison of point estimates
(Asymptotic data set, E-B approximation)

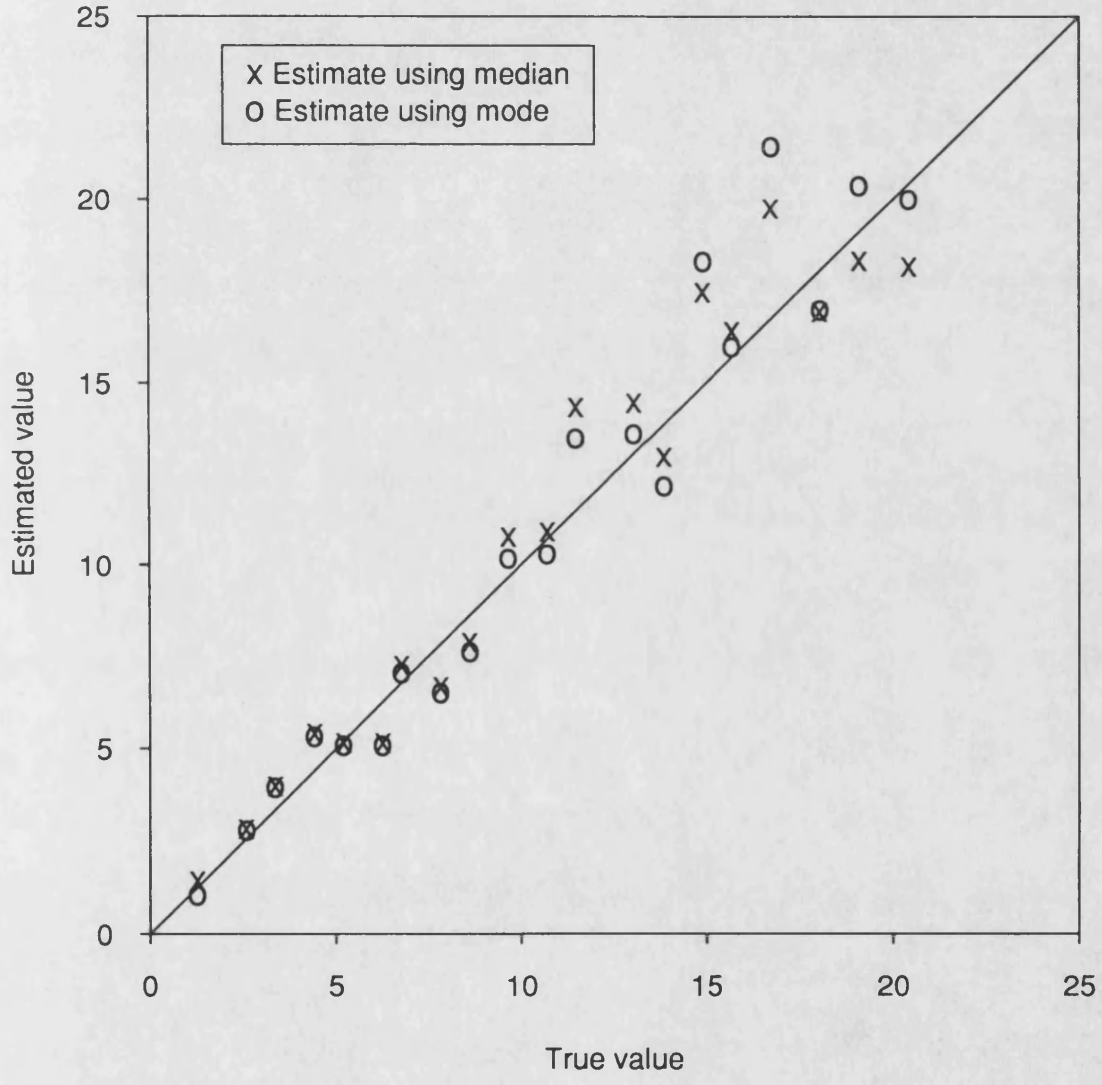


Table 5.5 Mean absolute error when using the mode and a triangular prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Gompertz	0.45	0.45	0 - 13.43
Preece-Baines	0.14	0.14	4 - 19.8
Bleasdale-Nelder	3.26	3.25	20 - 178

Table 5.6 Mean absolute error when using the median and a triangular prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Gompertz	0.43	0.43	0 - 13.43
Preece-Baines	0.14	0.14	4 - 19.8
Bleasdale-Nelder	3.22	3.21	20 - 178

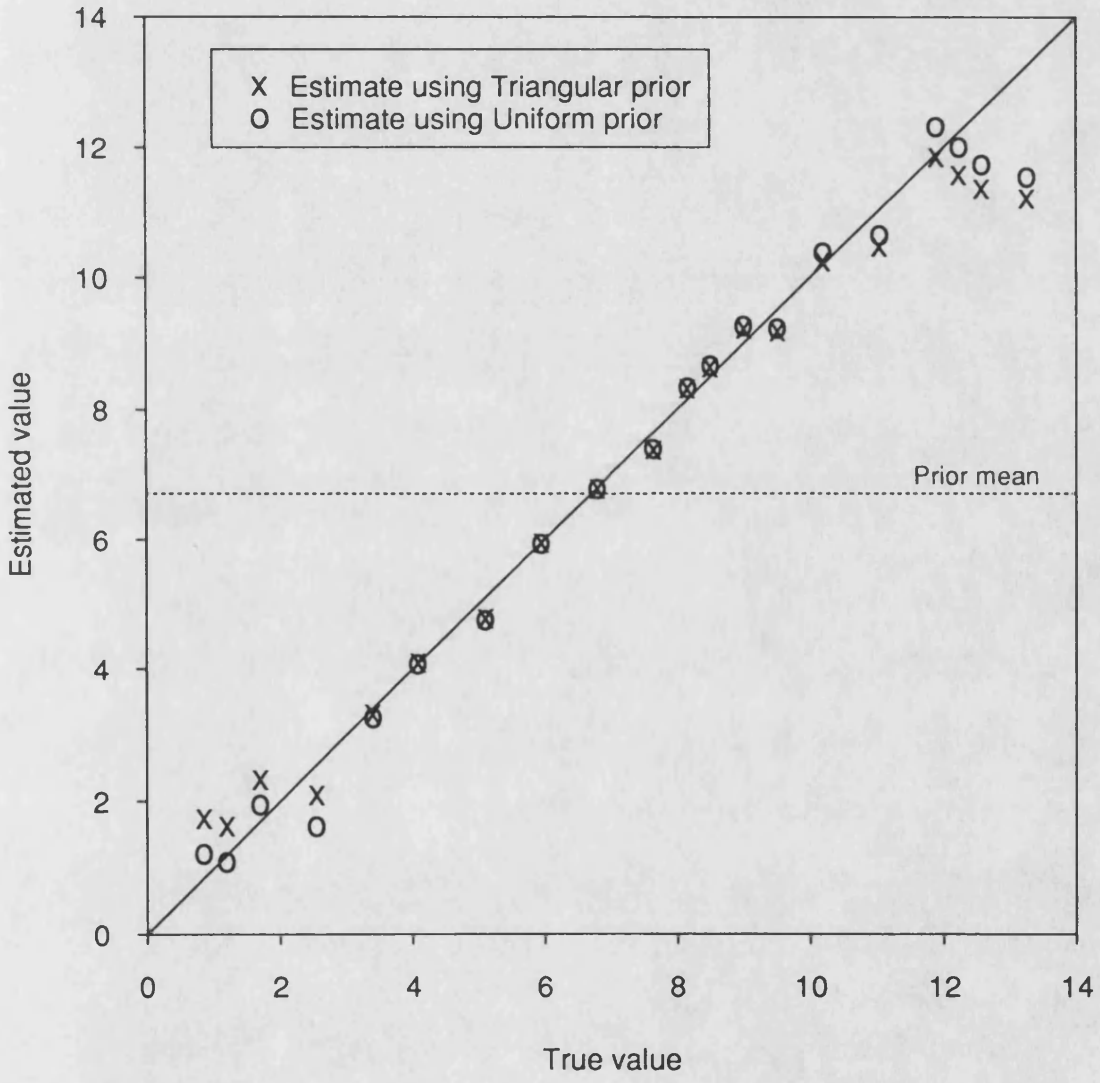
Table 5.7 Mean absolute error when using the mode and a Student prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Weibull	0.74	0.74	10 - 79.52
Asymptotic	1.26	1.25	1 - 21.74

Table 5.8 Mean absolute error when using the median and a Student prior for ξ

Data set	E-B approximation	S-S approximation	Calibration range
Weibull	0.75	0.75	10 - 79.52
Asymptotic	1.16	1.15	1 - 21.74

Fig.5.6: Comparison of point estimates
(Gompertz data set, E-B approximation)



Tables 5.9 and 5.10 give details of the symmetrical and HPD interval estimates for ξ respectively. Both the E-B and S-S approximations produce good interval estimates for ξ and comparison of the results in the two tables shows that those obtained using the S-S approximation are marginally more precise than those obtained using the E-B approximation. Where the posterior distributions of ξ are nearly symmetrical the HPD and symmetrical interval estimates for ξ will differ very little but where the distributions are skewed, the HPD interval will by definition give the shortest interval for a given probability content. Many of the posterior distributions of ξ were skewed and this is reflected in Tables 5.9 and 5.10 where the mean width of the HPD interval estimates for ξ is smaller than that of the symmetrical interval estimates for all the data sets and with both approximations.

Tables 5.11 and 5.12 give details of the interval estimates for ξ when using a triangular prior for ξ and a truncated Student prior for ξ respectively. Here only symmetrical interval estimates were constructed. Examination of these results shows that the differences in precision of interval estimates obtained using the two approximations is negligible and in fact the E-B approximation has a higher proportion of successful estimates for the Bleasdale-Nelder data set when using a triangular prior. Comparison of Tables 5.11 and 5.12 with the corresponding results in Table 5.9 shows that as expected interval estimates obtained when using a triangular or Student prior are generally narrower than those obtained when using a uniform prior.

5.7.2 Comparison of Imhof's method, the Eagleson-Buckley approximation and the Solomon-Stephens approximation

The posterior density $\pi(\xi|\mathbf{Y}',\mathbf{z})$ using Imhof's method is given by equation (5.3.6) and is very complicated because it involves a double integral and one of the integrands $\cos\theta(u,y)/\rho(u)$ is oscillatory (see Fig. 5.1). In addition $\theta(u,y)$ and $\rho(u)$ are defined as

$$\theta(u,y) = \tfrac{1}{2}\sum_{i=3}^{n} \tan^{-1}(\lambda_{iD}u) - \frac{(RSS)\,uy}{E_{\xi}^{*}} \qquad (5.7.1)$$

$$\rho(u) = \prod_{i=3}^{n} (1+\lambda_{iD}^{2}\,u^{2}). \qquad (5.7.2)$$

Equations (5.7.1) and (5.7.2) are fairly time-consuming to evaluate when $n = 80$ as it is for the five simulated data sets.

Table 5.9 Symmetrical interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Data set		Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Gompertz	E-B	2.84	0.82	1.82	19/20
	S-S	2.83	0.81	1.80	19/20
Weibull	E-B	10.15	2.51	4.59	20/20
	S-S	9.97	2.55	4.57	20/20
Preece-Baines	E-B	3.40	0.24	0.94	19/20
	S-S	3.40	0.23	0.93	19/20
Bleasdale-Nelder	E-B	32.54	4.84	16.51	18/20
	S-S	31.97	4.75	16.17	18/20
Asymptotic	E-B	9.54	1.43	5.84	20/20
	S-S	9.46	1.41	5.76	20/20

Table 5.10 HPD interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Data set		Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Gompertz	E-B	2.62	0.82	1.79	19/20
	S-S	2.82	0.80	1.77	19/20
Weibull	E-B	10.06	2.51	4.55	20/20
	S-S	9.83	2.44	4.42	20/20
Preece-Baines	E-B	3.28	0.20	0.93	19/20
	S-S	3.06	0.19	0.90	19/20
Bleasdale-Nelder	E-B	32.54	4.83	16.47	18/20
	S-S	31.93	4.74	16.14	18/20
Asymptotic	E-B	9.55	1.25	5.69	19/20
	S-S	9.41	1.23	5.59	19/20

Table 5.11 Interval estimates for ξ assuming a triangular prior for ξ (90% probability level)

Data set		Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Gompertz	E-B	2.29	0.82	1.67	19/20
	S-S	2.26	0.81	1.65	19/20
Preece-Baines	E-B	2.60	0.29	0.87	19/20
	S-S	2.57	0.28	0.87	19/20
Bleasdale-Nelder	E-B	28.22	4.58	15.66	18/20
	S-S	27.84	4.50	15.34	17/20

Table 5.12 Interval estimates for ξ assuming a Student prior for ξ (90% probability level)

Data set		Maximum interval width	Minimum interval width	Mean interval width	Proportion of successful estimates
Weibull	E-B	9.69	2.51	4.53	20/20
	S-S	9.53	2.55	4.52	20/20
Asymptotic	E-B	8.58	1.52	5.67	20/20
	S-S	8.52	1.49	5.59	20/20

The evaluation of the normalising constant C_I in equation (5.3.6) involves triple numerical integration. The result of all these facts is that the evaluation of the posterior distribution $\pi(\xi \mid \mathbf{Y}', \mathbf{z})$ is very time-consuming. For the assessment of the E-B and S-S approximations twenty values of X were chosen to cover the calibration range and twenty posterior distribution of ξ constructed for each data set using a variety of prior densities for ξ. Because it was so time-consuming to construct a posterior distribution of ξ using Imhof's method, it was decided to assume a uniform prior *only* for ξ and construct the twenty posterior distributions for the Bleasdale-Nelder data set *only*. Fig. 5.7 shows two posterior density functions, $\pi(\xi|\mathbf{Y}', \mathbf{z})$ for the Bleasdale-Nelder data set using Imhof's method and the S-S approximation. The true value of ξ is 162.0. The differences between the two density functions are very small. Fig. 5.8 shows two posterior density functions, $\pi(\xi|\mathbf{Y}', \mathbf{z})$ for the Bleasdale-Nelder data set using Imhof's method and the E-B approximation. The true value of ξ is 96.0. This shows the two posterior densities to be identical. Fig. 5.9 shows the twenty posterior medians obtained using Imhof's method, the E-B approximation and the S-S approximation, plotted against the true value ξ. The 45 ° line represents zero error. The figure shows that the point estimates obtainable from the three approaches are virtually identical. The mean absolute errors are the same to one decimal place too, whether the mode is used or the median is used. Fig. 5.9 also shows that all three approaches produce good point estimates of ξ.

This degree of closeness in the median and modal estimates obtained from the three approaches, is mirrored in the remaining four data sets namely, Gompertz, Weibull, Asymptotic and Preece-Baines data sets. For these four data sets, three values of X were chosen, two from either end of the calibration range and one from the middle. Three posterior distributions of ξ were constructed for each data set, assuming a uniform prior for ξ. Tables 5.13 and 5.14 give the point estimates $\hat{\xi}_{\text{mode}}$ and $\hat{\xi}_{\text{median}}$ for the four data sets. In some cases, the point estimates of ξ obtained from the three approaches, are identical, whether one uses the posterior mode or the posterior median and in the majority of cases they agree to one decimal place.

Three symmetrical interval estimates for ξ and three HPD interval estimates for ξ were constructed using a 90% probability level. Tables 5.15 and 5.16 give the widths of these interval estimates. The *unsuccessful* interval estimates are denoted by (F). Study of these results reveals that there are minimal differences in the three interval estimates with respect to both width and number of successful estimates. The interval estimates obtained using the S-S approximation appear to be marginally more precise.

Fig.5.7: Comparison of posterior densities ($\xi = 162.0$)

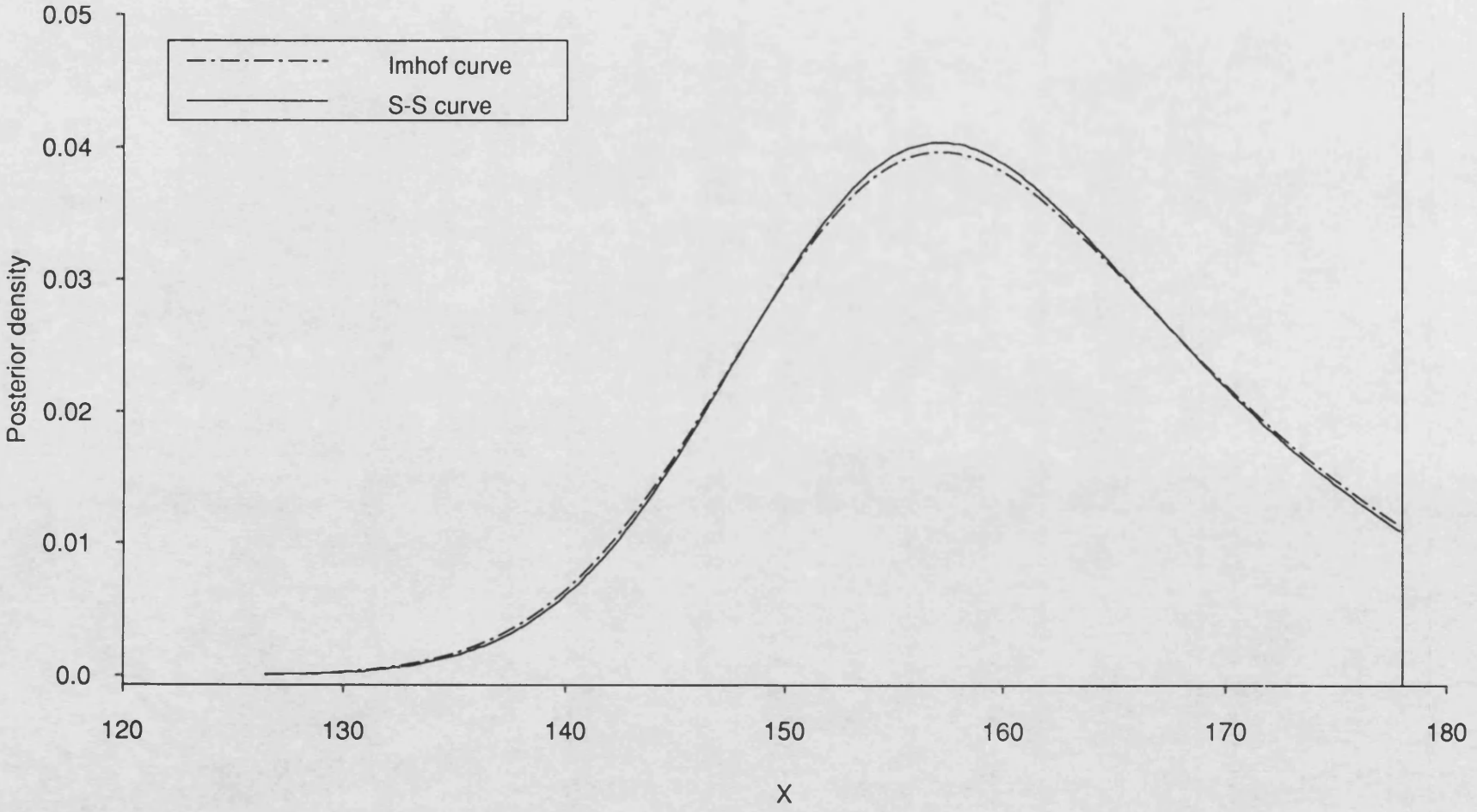


Fig.5.8: Comparison of posterior densities ($\xi = 96.0$)

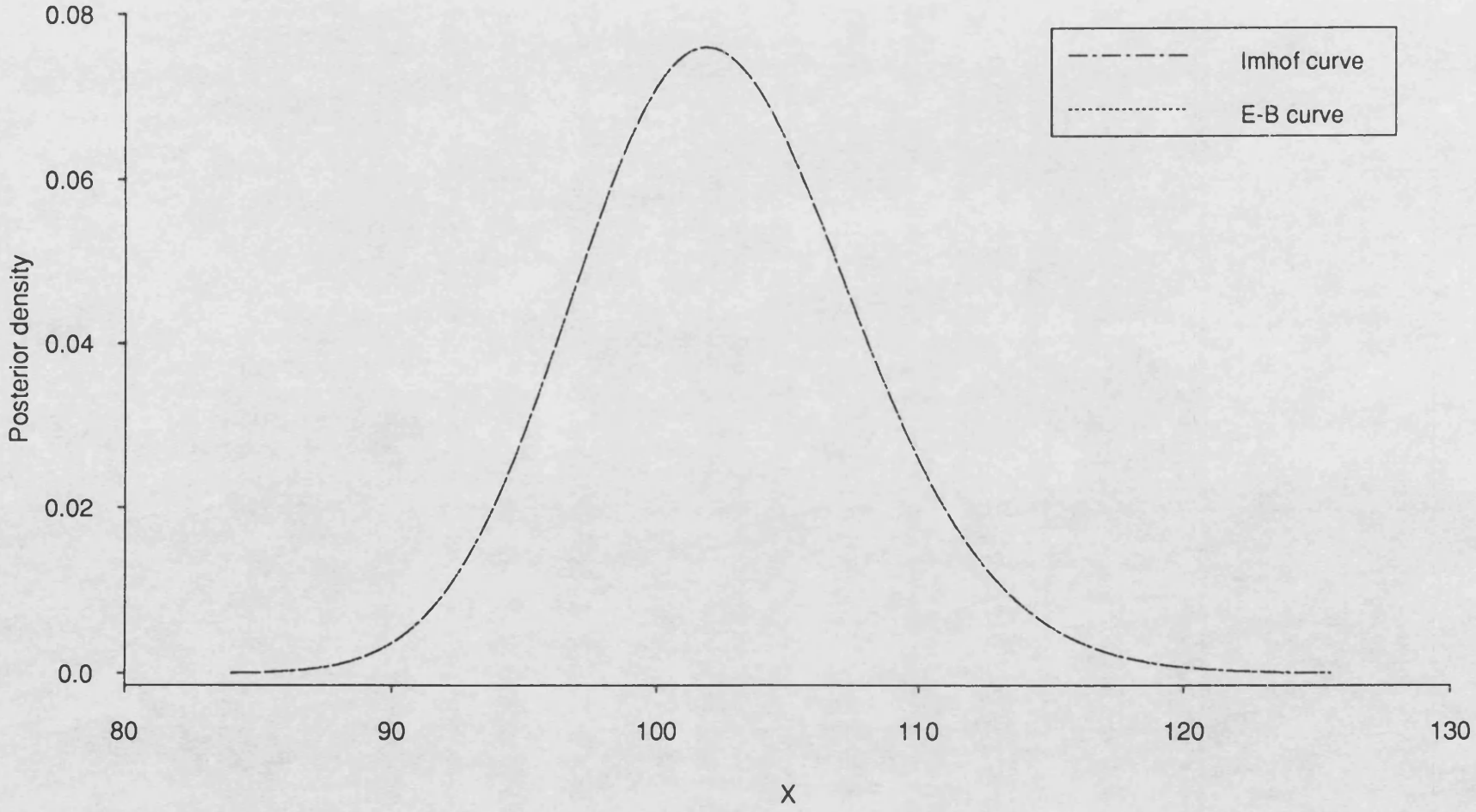


Fig.5.9: Comparison of point estimates

(Bleasdale-Nelder data set)

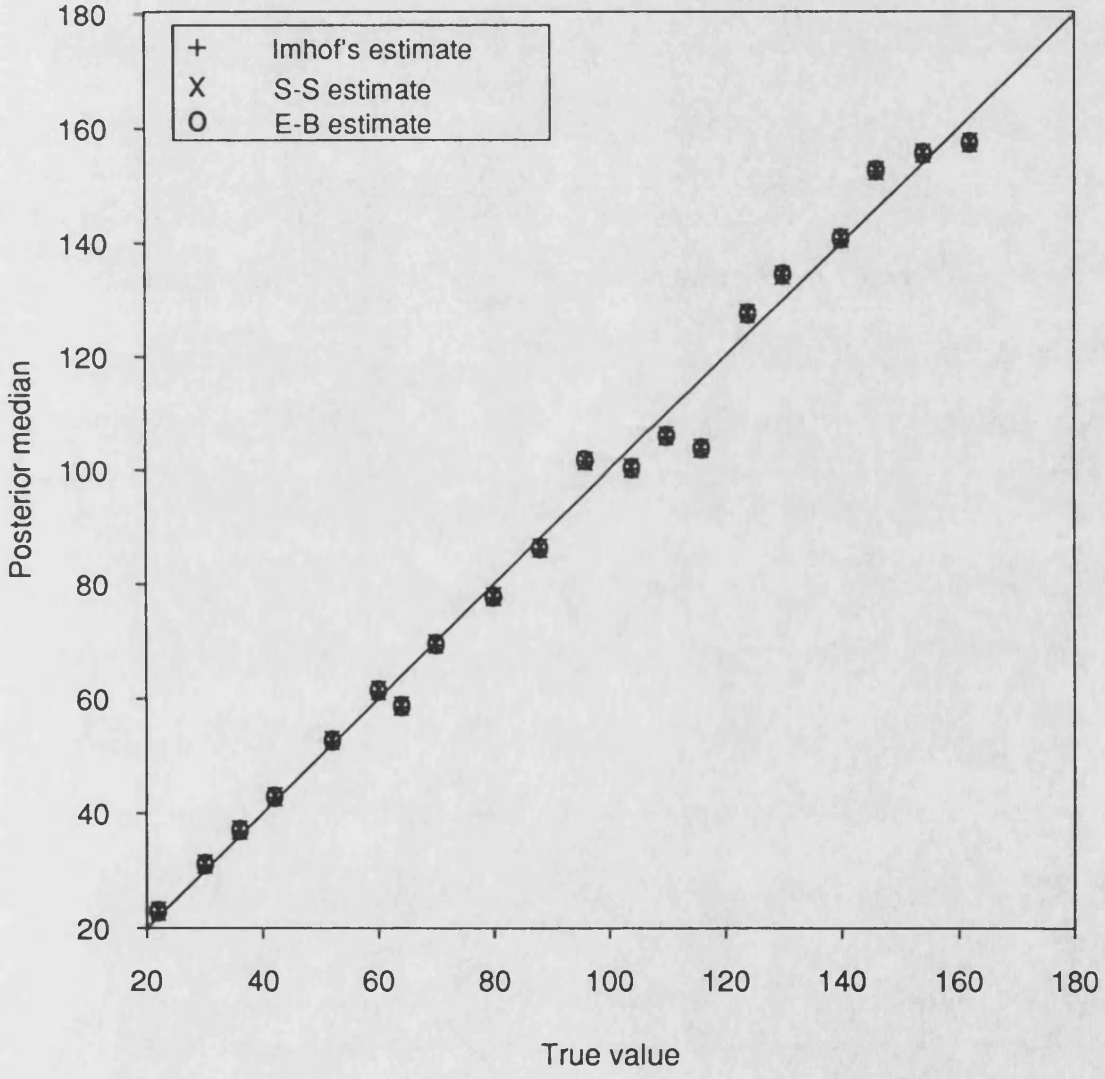


Table 5.13 Point estimate $\hat{\xi}_{mode}$ assuming a uniform prior for ξ

Date set	E-B approximation	S-S approximation	Imhof's method	True value of ξ
Gompertz	1.31	1.31	1.31	0.85
	8.37	8.37	8.37	8.16
	12.56	12.56	12.56	11.90
Weibull	11.98	11.98	11.94	11.75
	44.90	44.90	44.42	45.00
	75.44	75.45	75.22	76.50
Preece-Baines	4.00	4.00	4.00	4.00
	10.31	10.31	10.30	10.40
	17.94	17.94	17.96	17.80
Asymptotic	2.90	2.90	2.90	2.58
	10.40	10.40	10.40	10.71
	20.10	20.11	20.10	20.43

Table 5.14 Point estimate $\hat{\xi}_{median}$ assuming a uniform prior for ξ

Date set	E-B approximation	S-S approximation	Imhof's method	True value of ξ
Gompertz	1.27	1.26	1.27	0.85
	8.40	8.40	8.40	8.16
	12.38	12.39	12.38	11.90
Weibull	12.12	12.11	12.09	11.75
	44.91	44.92	45.00	45.00
	75.01	75.05	75.04	76.50
Preece-Baines	4.08	4.07	4.07	4.00
	10.31	10.31	10.31	10.40
	18.23	18.24	18.24	17.80
Asymptotic	2.93	2.93	2.93	2.58
	11.01	10.99	11.01	10.71
	18.25	18.28	18.25	20.43

Table 5.15 Widths of symmetrical interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Date set	E-B approximation	S-S approximation	Imhof's method	Calibration range
Gompertz	2.45 1.17 2.20	2.43 1.16 2.18	2.45 1.17 2.20	0 - 13.43
Weibull	3.99 2.77 8.93	3.91 2.77 8.79	3.75 3.63 8.83	10 - 79.52
Preece-Baines	0.24(F) 0.81 2.93	0.23(F) 0.80 2.88	0.24(F) 0.81 2.93	4 - 19.8
Bleasdale-Nelder	4.84 17.47 31.27	4.75 17.06 30.84	4.84 17.47 31.26	20 - 178
Asymptotic	2.48 7.42 7.55	2.44 7.25 7.47	2.48 7.42 7.55	1 - 21.74

Table 5.16 Widths of H.P.D. interval estimates for ξ assuming a uniform prior for ξ (90% probability level)

Date set	E-B approximation	S-S approximation	Imhof's method	Calibration range
Gompertz	2.35 1.18 2.06	2.34 1.15 2.04	2.35 1.17 2.06	0 - 13.43
Weibull	3.86 2.77 8.66	3.79 2.69 8.32	3.75 2.94 8.41	10 - 79.52
Preece-Baines	0.20 0.81 2.85	0.19 0.78 2.83	0.20 0.81 2.85	4 - 19.8
Bleasdale-Nelder	4.83 17.42 31.04	4.74 17.03 30.62	4.81 17.42 31.04	20 - 178
Asymptotic	2.48 7.26 7.13	2.44 7.03 7.06	2.47 7.26 7.19	1 - 21.74

It is interesting to note that the only unsuccessful interval estimate is that obtained for the Preece-Baines data set using a symmetrical interval estimate. The corresponding highest posterior density interval estimate is successful. This is because the posterior density function $\pi(\xi|\mathbf{Y}',\mathbf{Z})$ is skewed to the right (see Fig. 5.3). One would expect the posterior distribution of ξ to be positively or negatively skewed where the true value, ξ was very close to or equal to x_1 or x_n respectively since all the information pertaining to the value of ξ is contained within the calibration range $[x_1, x_n]$. Because of the positive skewness in this particular posterior distribution, both the posterior mode and HPD interval are better estimates for ξ than the posterior median and symmetrical interval estimate (see Tables 5.13, 5.14, 5.15 and 5.16).

5.8 CONCLUDING REMARKS

This chapter has considered four approaches based on predictive densities.

(a) Imhof's method (Section 5.3)

(b) Eagleson-Buckley approximation (Section 5.4)

(c) Solomon-Stephens approximation (Section 5.5)

(d) Interpolating spline approach (Section 5.6)

Approaches (a) - (c) involve an approximation to the distribution of the residual sum of squares associated with the calibration experiment and result in a predictive density function, $L(\xi)$, which does not have a closed analytical form but which requires numerical integration for its evaluation. Approach (d) results in a predictive density function which is a non-central Student density.

All four approaches (a) - (d) resulted in good point and interval estimates for ξ. Because many of the posterior distributions of ξ were skewed the highest posterior density (HPD) interval estimates for ξ were narrower than the corresponding symmetric interval estimates and in the case of the antibiotic assay data , the use of HPD interval estimates rather than symmetrical interval estimates resulted in a higher number of successful estimates. It would therefore seem sensible to use HPD interval estimates for ξ. With regard to point estimates, the differences between the posterior mode and posterior median were most pronounced when a uniform prior was used. In this case, the posterior mode seemed to be marginally better as a point estimate than the median. Because of the skewness of many of the posterior distributions of ξ it might be argued that the median should be used as it is a robust estimator. However the mode is the *only* point estimate which lies within the $100(1-\alpha)\%$ HPD interval for all α $(0 \leq \alpha \leq 1)$. It is therefore recommended that the posterior mode is used as a

point estimator of ξ, in conjunction with highest posterior density interval estimates for ξ.

With regard to the three approaches (a) - (c), the Solomon-Stephens approximation produced marginally more precise interval estimates for ξ. The Solomon-Stephens approximation utilised *all* the degrees of freedom associated with the residual sum of squares, *RSS*, whereas the Eagleson-Buckley approximation was based on a much lower number of degrees of freedom. Both the Eagleson-Buckley and Solomon-Stephens approximations worked well on all the data sets, were quick and easy to use and did not appear to have any disadvantages from either the statistical or computing point of view. However, Imhof's method was not easy to implement and was very time-consuming. Despite using a Nag routine ideally suited to an oscillatory integrand, there were problems with round-off error when evaluating $\int_0^{U^*} \frac{\cos\theta(u,y)}{\rho(u)}\,du$ where $\theta(u,y)$ and $\rho(u)$ are given by equations (5.7.1) and (5.7.2). The results obtained using Imhof's method were virtually identical to those obtained using the Eagleson-Buckley approximation, for all the data sets except the Weibull data set and for this latter data set the Solomon-Stephens approximation produced marginally better results overall than those obtained using Imhof's method.

In conclusion therefore, the following recommendations are made :

(i) Use the Solomon-Stephens approximation to the distribution of *RSS*.

(ii) Use the posterior mode as a point estimator of ξ.

(iii) Use highest posterior density intervals as interval estimates for ξ.

Recommendations (ii) and (iii) apply to approach (d) where the spline is an interpolating rather than a smoothing spline. This was the approach used for the antibiotic assay data.

All the results obtained in this chapter using the Solomon-Stephens approximation will be compared in Chapter 7 with the corresponding results in Chapters 3 and 4.

6. TWO NON-PARAMETRIC APPROACHES TO PROPORTIONS

6.1 INTRODUCTION

This chapter will consider firstly a non-parametric approach to proportion data arising from a Binomial model. Silverman (1985) states that the spline smoothing approach, which was first considered in Chapter 3, can be extended to produce *generalised smoothing* which has certain similarities with generalised linear models (McCullagh and Nelder, 1989). Consider the model given by

$$Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,n$$

where $a < x_1 < x_2,\ldots, < x_n < b$, $f \in W_2$ and the random errors ε_i $(i=1,2,\ldots,n)$ are assumed to be uncorrelated with mean zero and variance σ^2/w_i where the weights are known. In Section 3.1.2, Chapter 3, the following constrained optimisation problem was considered : minimise

$$\sum_{i=1}^{n} w_i (f(x_i) - Y_i)^2 + \alpha \int_a^b [f''(x)]^2 dx \qquad (6.1.1)$$

subject to $f \in W_2$ and $\alpha > 0$ (i.e. expression (3.1.9) with $m = 2$). The solution is a natural cubic smoothing spline with knots at the data points $\{ x_i \}$ $i=1,2,3,\ldots,n$.

Suppose now the sum of squares term in (6.1.1) is replaced by a function $\rho(Y_i, f(x_i))$ where $\rho(\mathbf{Y}, \varphi)$ is a function of a real parameter φ and the observation $\mathbf{Y}$. It is usual to take $-\frac{1}{2}\rho(\mathbf{Y},\varphi)$ as the log likelihood or partial likelihood of φ given $\mathbf{Y}$ for some parametric family of distributions. In this thesis, $-\frac{1}{2}\rho(\mathbf{Y},\varphi)$ will be taken to be the log likelihood of φ given $\mathbf{Y}$. Let

$$S_{GL} = \sum_{i=1}^{n} \rho(Y_i, f(x_i)) + \alpha \int_a^b [f''(x)]^2 dx \qquad (6.1.2)$$

Now $-\frac{1}{2}S_{GL}(f)$ is a penalised version of the log likelihood function of φ given $\mathbf{Y}$ where $\varphi^{\mathrm{T}} = (f(x_1) f(x_2),\ldots,f(x_n))$ and varying the choice of ρ produces non-parametric versions of many regression techniques, such as logistic regression (Silverman, 1985). As mentioned above, the first approach of the chapter is concerned with Binomial proportion data so the approach produces a non-parametric version of logistic regression.

6.2 A NON-PARAMETRIC VERSION OF LOGISTIC REGRESSION

Suppose that N individuals/experimental units are under study and that for each individual/unit there are only two responses, a positive response and a negative response. We will assume that there is only one explanatory variable, X. Suppose also that these N individuals/units are divided into n homogeneous groups of size m_i (i=1,2,3,...,n) such that the ith group, of size m_i, corresponds to level i of X, x_i. Therefore $m_1 + m_2 + m_3 + \ldots + m_n = N$. We are interested in observations of the form $\frac{Z_1}{m_1}, \frac{Z_2}{m_2}, \ldots, \frac{Z_n}{m_n}$ where Z_i is the number of positive responses within the ith group (i=1,2,3,...,n). If all the observations on the individuals/units within the *same* group are independent and P(A positive response) = θ_i is the *same* for all individuals/units within the *same* group, then Z_i is a Binomial random variable with parameters m_i and θ_i (i=1,2,...,n). The log likelihood function of $\boldsymbol{\theta}$ given $\mathbf{Z}$ is given by

$$\ln \mathcal{L}(\boldsymbol{\theta}\,;\mathbf{Z}) = \sum_{i=1}^{n} Z_i \ln\left[\frac{\theta_i}{1-\theta_i}\right] + \sum_{i=1}^{n} m_i \ln(1-\theta_i)$$

where $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1\, \theta_2\, \theta_3 \ldots\ldots\, \theta_n)$. Let $Y_i = Z_i/m_i$ then the log likelihood function is given by

$$\begin{aligned}\ln \mathcal{L}(\boldsymbol{\theta}\,;\mathbf{Y}) &= \sum_{i=1}^{n} m_i Y_i \ln\left[\frac{\theta_i}{1-\theta_i}\right] + \sum_{i=1}^{n} m_i \ln(1-\theta_i) \\ &= \sum_{i=1}^{n} m_i \left[Y_i \ln\left[\frac{\theta_i}{1-\theta_i}\right] + \ln(1-\theta_i)\right]. \qquad (6.2.1)\end{aligned}$$

The logistic model assumed for the data is given by

$$logit(\theta_i) = \ln\left[\frac{\theta_i}{1-\theta_i}\right] = f(x_i) \qquad (6.2.2)$$

where f is a smooth function. An equivalent statement of model (6.2.2) is given by

$$E\left[\frac{Z_i}{m_i}\right] = E(Y_i) = \theta_i = \frac{e^{f(x_i)}}{1+e^{f(x_i)}} \qquad i=1,2,\ldots,n. \qquad (6.2.3)$$

It was stated above that $-\frac{1}{2}\rho(\mathbf{Y},\boldsymbol{\varphi})$ is taken to be the log likelihood, so $\rho(Y_i\,,\varphi_i)$ is defined by

$$\rho(Y_i,\varphi_i) = -2m_i\left[Y_i \ln\left[\frac{\theta_i}{1-\theta_i}\right] + \ln(1-\theta_i)\right]$$

where $\theta_i = e^{\varphi_i}/(1+e^{\varphi_i})$. Using equation (6.2.3) and the fact that $\varphi_i = f(x_i)$

gives

$$\rho(Y_i, f(x_i)) = -2m_i\left[Y_i f(x_i) + \ln\left[\frac{e^{-f(x_i)}}{1+e^{-f(x_i)}}\right]\right]$$

$$= 2m_i[(1-Y_i)f(x_i) + \ln(1+e^{-f(x_i)})]. \qquad (6.2.4)$$

Substituting for $\rho(Y_i, f(x_i))$ in equation (6.1.2) gives the following expression for S_{GL},

$$S_{GL}(f) = 2\sum_{i=1}^{n} m_i[(1-Y_i)f(x_i) + \ln(1+e^{-f(x_i)})]$$

$$+ \alpha\int_a^b [f''(x)]^2 dx. \qquad (6.2.5)$$

Suppose that $S_{GL}(f)$ has a finite minimum at $f = \hat{f}$, it can be shown that $\hat{f}$ is a spline function satisfying (3.1.4), with $p = 4$, (Silverman, 1985). So $\hat{f}$ is a natural cubic spline with knots $\{x_i\}$. Silverman (1985) states that $\hat{f}$ will exist provided that $\sum_{i=1}^{n} m_i[(1-Y_i)f(x_i) + \ln(1+e^{-f(x_i)})]$ has a finite minimum in the space of all *linear* functions f.

6.2.1 A quadratic approximation to the log likelihood

Let $\mathbf{f}^T = (f(x_1)f(x_2)\ldots f(x_n))$. From (6.1.2)

$$S_{GL} = \sum_{i=1}^{n} \rho(Y_i, f(x_i)) + \alpha\int_a^b [f''(x)]^2 dx.$$

It will be recalled that $-\frac{1}{2}\rho(\mathbf{Y}, \mathbf{f})$ is the log likelihood of $\mathbf{f}$ given $\mathbf{Y}$, so let $-\frac{1}{2}\rho(\mathbf{Y}, \mathbf{f}) = \ln L(\mathbf{f}, \mathbf{Y})$. Then,

$$-\tfrac{1}{2}S_{GL} = \ln L(\mathbf{f}, \mathbf{Y}) - \tfrac{1}{2}\alpha\int_a^b [f''(x)]^2 dx. \qquad (6.2.6)$$

Suppose, as in Section 3.2.2, Chapter 3, a prior for $\mathbf{f}$ of the form

$$p(\mathbf{f}) \propto \frac{-\alpha}{2\sigma^2}\int_a^b [f''(x)]^2 dx$$

is assumed. If one takes $\sigma^2 = 1$, then $-\frac{1}{2}S_{GL}$, given by equation (6.2.6), represents the posterior log likelihood of $\mathbf{f}$. Using the same notation as Section 3.1.3, Chapter 3,

$$\mathbf{f} = B\boldsymbol{\gamma} \text{ and } \int_a^b [f''(x)]^2 dx = \boldsymbol{\gamma}^T\Omega\boldsymbol{\gamma}.$$

The n x n matrix Ω is a symmetric positive semi-definite matrix. Substituting for

γ gives

$$\alpha \int_a^b [f''(x)]^2 dx = \alpha (B^{-1}\mathbf{f})^{\mathrm{T}} \Omega (B^{-1}\mathbf{f}) = \mathbf{f}^{\mathrm{T}} E \mathbf{f}$$

where $E = \alpha (B^{-1})^{\mathrm{T}} \Omega B^{-1} = \alpha (B^{\mathrm{T}})^{-1} \Omega B^{-1}$. Now from equation (3.1.13c), the spline smoother, $\hat{f}_\alpha$, which uniquely minimises (6.1.1) satisfies the equation

$$(B^{\mathrm{T}} W B + \alpha \Omega) B^{-1} \hat{\mathbf{f}}_\alpha = B^{\mathrm{T}} W \mathbf{Y}$$

Pre-multiplying the above by $(B^{\mathrm{T}})^{-1}$ gives

$$(W + E) \hat{\mathbf{f}}_\alpha = W \mathbf{Y}$$

where W is an n x n diagonal matrix with entries w_i and $\mathbf{Y}$ is an n x 1 vector given by $\mathbf{Y}^{\mathrm{T}} = (\bar{Y}_1 \bar{Y}_2 \ldots\ldots \bar{Y}_n)$. It follows that the n x n matrix E is independent of the choice of B-splines and it can be shown that E is a symmetric positive semi-definite matrix. So equation (6.2.6) becomes

$$-\tfrac{1}{2} S_{GL} = \ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y}) - \tfrac{1}{2} \mathbf{f}^{\mathrm{T}} E \mathbf{f}. \qquad (6.2.7)$$

Let $\hat{\mathbf{f}}^{\mathrm{T}} = (\hat{f}(x_1)\, \hat{f}(x_2)\, \hat{f}(x_3) \ldots\ldots \hat{f}(x_n))$ and let us consider expanding (up to the second-order) the log likelihood of $\mathbf{f}$ given $\mathbf{Y}$, $\ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y})$, about $\mathbf{f} = \hat{\mathbf{f}}$.

$$\ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y}) \simeq \ln \mathcal{L}(\hat{\mathbf{f}}\,;\mathbf{Y}) + (\mathbf{f}-\hat{\mathbf{f}}) \nabla \ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y})\big|_{\mathbf{f}=\hat{\mathbf{f}}} - \tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}} K (\mathbf{f}-\hat{\mathbf{f}}). \qquad (6.2.8)$$

The matrix K is called the sample information matrix (Edwards, 1972) and it fully specifies the quadratic approximation to the log likelihood surface in the region of $\mathbf{f} = \hat{\mathbf{f}}$. The ijth element of K is defined as

$$K_{ij} = \frac{-\partial^2 \ln \mathcal{L}(\mathbf{f}(x))}{\partial f_i \partial f_j}\Big|_{\mathbf{f}=\hat{\mathbf{f}}}$$

where $f_i = f(x_i)$ and $f_j = f(x_j)$.

$$\ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y}) \propto (\mathbf{f}-\hat{\mathbf{f}}) \nabla \ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y})\big|_{\mathbf{f}=\hat{\mathbf{f}}} - \tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}} K (\mathbf{f}-\hat{\mathbf{f}})$$

where $\propto$ means "as far as it involves $\mathbf{f}$ ". Substituting for $\ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y})$ in (6.2.7) gives approximately

$$-\tfrac{1}{2} S_{GL} \propto (\mathbf{f}-\hat{\mathbf{f}}) \nabla \ln \mathcal{L}(\mathbf{f}\,;\mathbf{Y})\big|_{\mathbf{f}=\hat{\mathbf{f}}} - \tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}} K (\mathbf{f}-\hat{\mathbf{f}}) - \tfrac{1}{2} \mathbf{f}^{\mathrm{T}} E \mathbf{f}. \qquad (6.2.9)$$

Suppose $\mathbf{V}$ has a multivariate normal distribution with mean μ, covariance matrix Σ and p.d.f $g(\mathbf{v})$. Then the log likelihood of $\mathbf{V}$ is given by

$$\ln g(\mathbf{v}) \propto -\tfrac{1}{2}(\mathbf{v}-\mu)^{\mathrm{T}} \Sigma^{-1} (\mathbf{v}-\mu) \qquad (6.2.10)$$

Since $-\frac{1}{2} S_{GL}$ represents the posterior log likelihood of $\mathbf{f}$, it can be seen by comparing (6.2.9) and (6.2.10) that the posterior distribution of f is an

approximate multivariate normal distribution.

6.2.2 The equivalent weighted least squares problem

Now let us look at the first two terms of (6.2.9) in more detail.

Consider firstly the term $(\mathbf{f}-\hat{\mathbf{f}})\nabla \ln \mathcal{L}(\mathbf{f};\mathbf{Y})|_{\mathbf{f}=\hat{\mathbf{f}}}$.

$$\nabla \ln \mathcal{L}(\mathbf{f};\mathbf{Y}) = \begin{bmatrix} \frac{\partial(\ln \mathcal{L})}{\partial f_1} \\ \frac{\partial(\ln \mathcal{L})}{\partial f_2} \\ \vdots \\ \frac{\partial(\ln \mathcal{L})}{\partial f_n} \end{bmatrix} = -\tfrac{1}{2} \begin{bmatrix} \frac{\partial \rho(Y_1,f_1)}{\partial f_1} \\ \frac{\partial \rho(Y_2,f_2)}{\partial f_2} \\ \vdots \\ \frac{\partial \rho(Y_n,f_n)}{\partial f_n} \end{bmatrix} = -\tfrac{1}{2}\rho_2(\mathbf{Y},\mathbf{f})$$

So the first term of (6.2.9) is given by

$$(\mathbf{f}-\hat{\mathbf{f}})\nabla \ln \mathcal{L}(\mathbf{f};\mathbf{Y})|_{\mathbf{f}=\hat{\mathbf{f}}} = -\tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})\,\rho_2(\mathbf{Y},\hat{\mathbf{f}}).$$

Consider the second term of (6.2.9) in more detail. In particular, consider a diagonal element of the sample information matrix K, K_{ii}.

$$K_{ii} = \frac{-\partial^2 \ln \mathcal{L}}{\partial f_i^2}\Big|_{f_i=\hat{f}(x_i)} = \tfrac{1}{2}\frac{\partial^2 \rho}{\partial f_i^2}\Big|_{f_i=\hat{f}(x_i)} = \tfrac{1}{2}\,\rho_{22}(Y_i,\hat{f}_i)$$

Now $\rho(Y_i,f_i)$ is a function of Y_i and $f(x_i)$ *only*, so $\dfrac{\partial \rho(Y_i,f_i)}{\partial f_j} = 0$. It therefore follows that $K_{ij} = 0$ $i \neq j$ and so K is a diagonal positive definite matrix with diagonal elements $\frac{1}{2}\rho_{22}(Y_i,\hat{f}_i)$. Hence the second term of (6.2.9) is given by

$$-\tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}} K\,(\mathbf{f}-\hat{\mathbf{f}}) = -\tfrac{1}{4}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}}\rho_{22}(\mathbf{Y},\hat{\mathbf{f}})(\mathbf{f}-\hat{\mathbf{f}}).$$

The expression (6.2.9) therefore becomes

$$-\tfrac{1}{2}S_{GL}(f) \propto -\tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})\rho_2(\mathbf{Y},\hat{\mathbf{f}}) - \tfrac{1}{4}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}}\rho_{22}(\mathbf{Y},\hat{\mathbf{f}})(\mathbf{f}-\hat{\mathbf{f}}) - \frac{\alpha}{2}\int_a^b [f''(x)]^2 dx$$

$$S_{GL}(f) \propto (\mathbf{f}-\hat{\mathbf{f}})\rho_2(\mathbf{Y},\hat{\mathbf{f}}) + \tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}}\rho_{22}(\mathbf{Y},\hat{\mathbf{f}})(\mathbf{f}-\hat{\mathbf{f}}) + \alpha\int_a^b [f''(x)]^2 dx.$$

Let $$w_i^* = \tfrac{1}{2}\rho_{22}(Y_i,f_i)|_{f_i=\hat{f}_i} = \tfrac{1}{2}\rho_{22}(Y_i,\hat{f}(x_i)) \qquad (6.2.11)$$

Now $$(\mathbf{f}-\hat{\mathbf{f}})\rho_2(\mathbf{Y},\hat{\mathbf{f}}) + \tfrac{1}{2}(\mathbf{f}-\hat{\mathbf{f}})^{\mathrm{T}}\rho_{22}(\mathbf{Y},\hat{\mathbf{f}})(\mathbf{f}-\hat{\mathbf{f}})$$

$$= \sum_{i=1}^{n}(f_i-\hat{f}_i)\,\rho_2(Y_i,\hat{f}_i) + \tfrac{1}{2}\sum_{i=1}^{n} w_i^*\,(f_i-\hat{f}_i)^2$$

Substituting into the above expression for S_{GL} gives

$$S_{GL}(f) \propto \sum_{i=1}^{n}(f_i-\hat{f}_i)\,\rho_2(Y_i,\hat{f}_i) + \sum_{i=1}^{n} w_i^*(f_i-\hat{f}_i)^2 + \alpha\int_a^b[f''(x)]^2dx. \tag{6.2.12}$$

Now
$$\sum_{i=1}^{n}(f_i-\hat{f}_i)\rho_2(Y_i,\hat{f}_i) + \sum_{i=1}^{n} w_i^*(f_i-\hat{f}_i)^2$$

$$= \sum_{i=1}^{n} w_i^*\left[\frac{\rho_2 f_i}{w_i^*} + f_i^2 - 2f_i\hat{f}_i + \hat{f}_i^2\right] - \sum_{i=1}^{n}\rho_2\hat{f}_i$$

$$= \sum_{i=1}^{n} w_i^*\left[f_i^2 - 2f_i\left[\hat{f}_i - \frac{\rho_2}{2w_i^*}\right] + \hat{f}_i^2\right] - \sum_{i=1}^{n}\rho_2\hat{f}_i\,. \tag{6.2.13}$$

Let
$$Y_i^* = \hat{f}_i - \frac{\rho_2}{2w_i^*} = \hat{f}(x_i) - \frac{\rho_2(Y_i,\hat{f}(x_i))}{2w_i^*} \tag{6.2.14}$$

then expression (6.2.13) becomes

$$\sum_{i=1}^{n} w_i^*(f(x_i)-Y_i^*)^2 - \sum_{i=1}^{n}\left[w_i^*\left(Y_i^{*2}-\hat{f}_i^2\right) + \rho_2\hat{f}_i\right].$$

Substituting into (6.2.12) gives

$$S_{GL} \propto \sum_{i=1}^{n} w_i^*(f(x_i)-Y_i^*)^2 + \alpha\int_a^b[f''(x)]^2dx \tag{6.2.15}$$

where $\propto$ means "as far as it involves f ". The values Y_i^* are called *pseudo-observations* and the weights w_i^* are called *pseudo-weights*.

The posterior distribution of f for the weighted least squares problem corresponding to expression (6.2.15) is precisely the approximate multivariate normal distribution referred to in Section 6.2.1 above. Hence all the methods for generating the posterior distribution of ξ described in Chapters 3, 4 and 5 can be applied here using the observations (x_i,Y_i^*) and associated weights w_i^* $(i=1,2,3\ldots,n)$. The first step in the procedure for generating the posterior distribution of ξ is to obtain the spline estimate $\hat{f}$, which uniquely minimises $S_{GL}(f)$ defined by equation (6.2.5).

6.2.3 The calculation of the spline estimate $\hat{\mathbf{f}}$

From equations (6.2.11) and (6.2.14) respectively, w_i^* and Y_i^* are defined as

$$w_i^* = \tfrac{1}{2}\rho_{22}(Y_i,\hat{f}(x_i))$$

$$Y_i^* = \hat{f}(x_i) - \frac{\rho_2(Y_i,\hat{f}(x_i))}{2w_i^*}$$

For the model under consideration, ρ is given by equation (6.2.4), namely,

$$\rho(Y_i, f(x_i)) = 2m_i[(1-Y_i)f(x_i) + \ln(1+e^{-f(x_i)})]$$

Let $u_i = f(x_i)$ then $\rho(Y_i, u_i)$ is given by

$$\rho(Y_i, u_i) = 2m_i[(1-Y_i)u_i + \ln(1+e^{-u_i})]$$

$$\rho_2(Y_i, u_i) = \frac{\partial \rho}{\partial u_i} = 2m_i[(1-Y_i) - \frac{1}{(1+e^{u_i})}]$$

$$\rho_{22}(Y_i, u_i) = \frac{\partial^2 \rho}{\partial u_i^2} = \frac{2m_i e^{u_i}}{(1+e^{u_i})^2}$$

So the equations for w_i^* and Y_i^* are given by

$$w_i^* = \frac{m_i e^{\hat{f}(x_i)}}{(1+e^{\hat{f}(x_i)})^2} \tag{6.2.16}$$

$$Y_i^* = \hat{f}(x_i) - \frac{m_i}{w_i^*}\left[(1-Y_i) - \frac{1}{1+e^{\hat{f}(x_i)}}\right] \tag{6.2.17}$$

The spline function which minimises the objective function (6.2.15) subject to $f \in W_2$ and $\alpha > 0$ can only be obtained by using an iterative process because as can be seen from equations (6.2.16) and (6.2.17), both Y_i^* and w_i^* depend on the fitted values, $\hat{f}(x_i)$, for which only current estimates are available. The process consists of the following steps:

(1) Fit a cubic smoothing spline to the observations (x_i, Y_i) using weights $w_i = 1.0$. This produces a spline estimate $\hat{f}_1$.

(2) Calculate the pseudo-observations Y_i^* and w_i^* using equations (6.2.16) and (6.2.17) with $\hat{f}(x_i) = \hat{f}_1(x_i)$.

(3) Fit a cubic smoothing spline to the observations (x_i, Y_i^*) using weights w_i^*. This produces a current spline estimate, $\hat{f}_c$.

(4) Calculate the pseudo-observations Y_i^* and w_i^* using equations (6.2.16) and (6.2.17) with $\hat{f}(x_i) = \hat{f}_c(x_i)$.

(5) Repeat steps (3) and (4) until the spline estimates converge.

6.2.4 Comparison with generalised linear models

The formulae for evaluating w_i^* and Y_i^* (equations (6.2.11) and (6.2.14) respectively) are closely related to those of generalised linear models (GLM, McCullagh and Nelder, 1989). Consider the equivalent generalised linear model, namely a linear logistic regression model given by

$$logit\ (\theta_i) = \ln\left[\frac{\theta_i}{1-\theta_i}\right] = \beta_0 + \beta_1 x_i \qquad i=1,2,\ldots,n.$$

The *linear predictor* is denoted by η_i and is defined as $\eta_i = \beta_0 + \beta_1 x_i$. Hence

$$\theta_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}.$$

Let $\mu_i = E(Z_i)$ where Z_i is the number of positive responses in the ith group of size m_i. The relationship between η_i and μ_i is given by the *link function*, g,

$$\eta_i = g(\mu_i) = \ln\left[\frac{\mu_i}{m_i-\mu_i}\right] \qquad i=1,2,\ldots,n.$$

With GLM, the maximum likelihood estimates of β_0 and β_1 are obtained by using iterative weighted least squares. The formulae equivalent to equations (6.2.11) and (6.2.14) respectively are given by

$$w_i^* = \frac{1}{V}\left[\frac{d\mu_i}{d\eta_i}\right]^2 \qquad i=1,2,\ldots,n$$

$$Y_i^* = \eta_i + (Z_i-\mu_i)\frac{d\eta_i}{d\mu_i} \qquad i=1,2,\ldots,n$$

where V is the variance function.

For the linear logistic model, these formulae become

$$w_i^* = m_i\theta_i(1-\theta_i)\big|_{\theta_i=\hat{\theta}_i}$$

$$= \frac{m_i e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}{[1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i}]^2} \qquad i=1,2\ldots\ldots n$$

$$Y_i^* = \left[\eta_i + \frac{(Z_i-m_i\theta_i)}{m_i\theta_i(1-\theta_i)}\right]_{\substack{\eta_i=\hat{\eta}_i \\ \theta_i=\hat{\theta}_i}}$$

$$= (\hat{\beta}_0+\hat{\beta}_1 x_i) + \frac{m_i}{w_i^*}\left[Y_i - \frac{e^{\hat{\beta}_0+\hat{\beta}_1 x_i}}{(1+e^{\hat{\beta}_0+\hat{\beta}_1 x_i})}\right] \qquad i=1,2,\ldots,n.$$

Comparison of the above equations with equations (6.2.16) and (6.2.17) reveal many similarities. The variable Y^* is called the *adjusted dependent variable* and the maximum likelihood estimates of β_0 and β_1 are obtained by using an iterative process which consists of regressing Y^* on X using weights $\frac{1}{V}\left[\frac{d\mu}{d\eta}\right]^2$, obtaining new estimates for β_0 and β_1 and continuing until changes in estimates are

sufficiently small.

6.2.5 Application to simulated data sets

Data sets were simulated from two sigmoidal growth models (Ratkowsky, 1983), in particular,

(1) **Richards model**

$$\theta = \frac{1}{[1+\exp(\beta-\gamma x)]^{1/\delta}}$$

$$f(x) = \ln\frac{\theta}{(1-\theta)} = \ln\left[\frac{1}{[1+\exp(\beta-\gamma x)]^{1/\delta}-1}\right] \qquad (6.2.18)$$

with $\beta = 5.691$, $\gamma = 0.777$ and $\delta = 1.619$.

(2) **Gompertz model**

$$\theta = \exp(-\exp(\beta-\gamma x))$$

$$f(x) = \ln\frac{\theta}{(1-\theta)} = \ln\left[\frac{1}{\exp(\exp(\beta-\gamma x))-1}\right] \qquad (6.2.19)$$

with $\beta = 2.106$ and $\gamma = 0.388$.

For each calibration data set there were sixty knots and four values of Y were simulated at each knot value. So the calibration experiment model is given by

$$Y_{ik} = \theta_i + \varepsilon_{ik} = \frac{e^{f(x_i)}}{1+e^{f(x_i)}} + \varepsilon_{ik} \qquad \begin{matrix} i=1,2,\ldots,60 \\ k=1,2,3,4 \end{matrix}$$

where θ_i and $f(x_i)$ for the Richards and Gompertz models are given by equations (6.2.18) and (6.2.19) respectively. It is assumed that $m_i Y_{ik}$ has a Binomial distribution with parameters m_i and θ_i. For both data sets $m_1 = m_2 = \ldots = m_{60} = 70$. The Gompertz and Richards calibration data sets therefore consisted of 240 observations (x_i, Y_{ik}) where Y_{ik} is the kth proportion of positive responses at knot value x_i. Using the iterative process outlined in Section 6.2.3 above, the spline estimate, $\hat{f}$, was calculated for each calibration data set. Suppose this is called the *iterative estimate.* Figs. 6.1 and 6.2 show the Richards model curve given by equation (6.2.18) and the iterative estimate, $\hat{f}$, using the original probability scale and the logit scale respectively. The spline estimate, $\hat{f}$, seems to be a good approximation to the model curve, particularly for $X > 4$.

Fig.6.1: Binomial proportions (Richards data set)

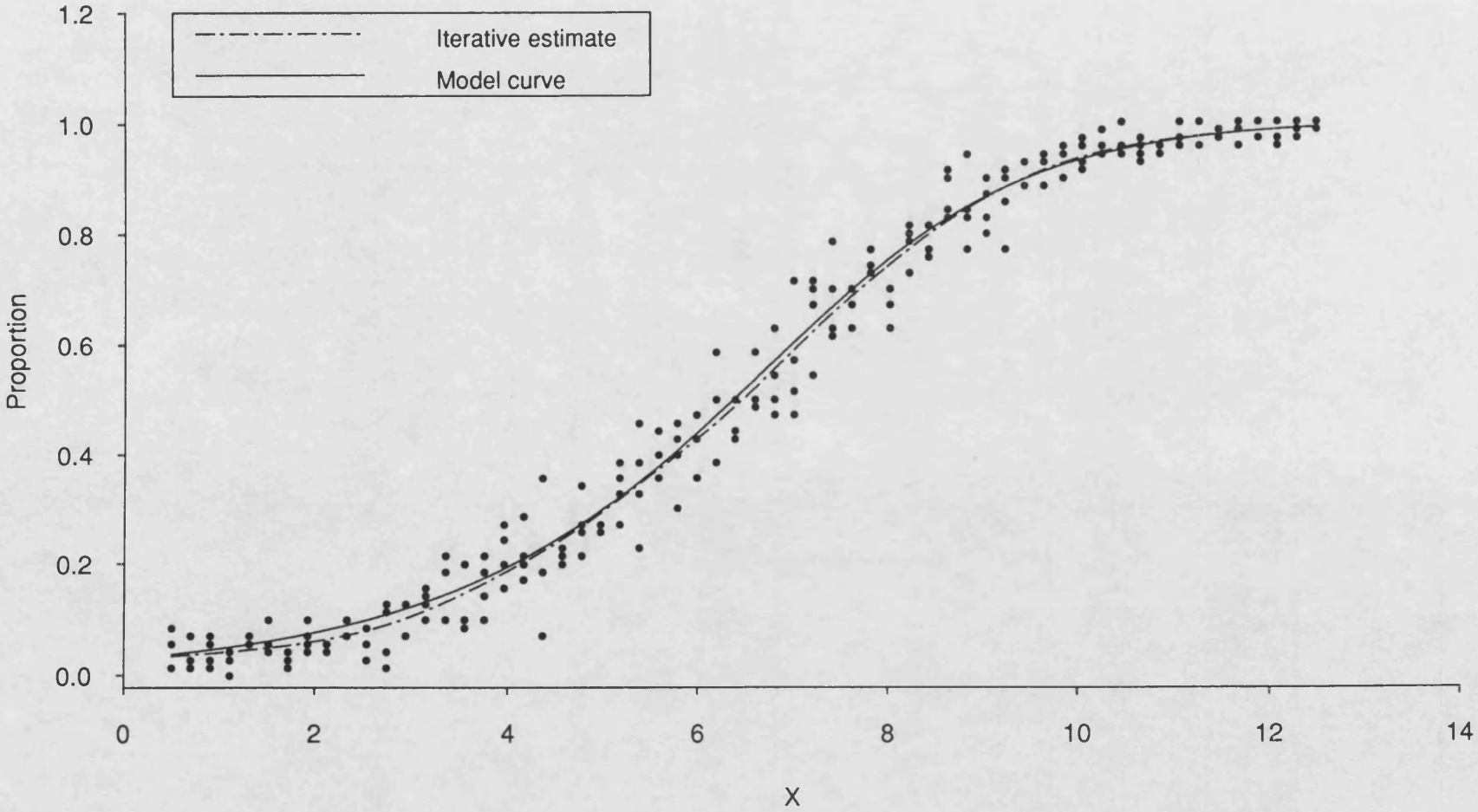


Fig.6.2: Transformed Binomial proportions (Richards data set)

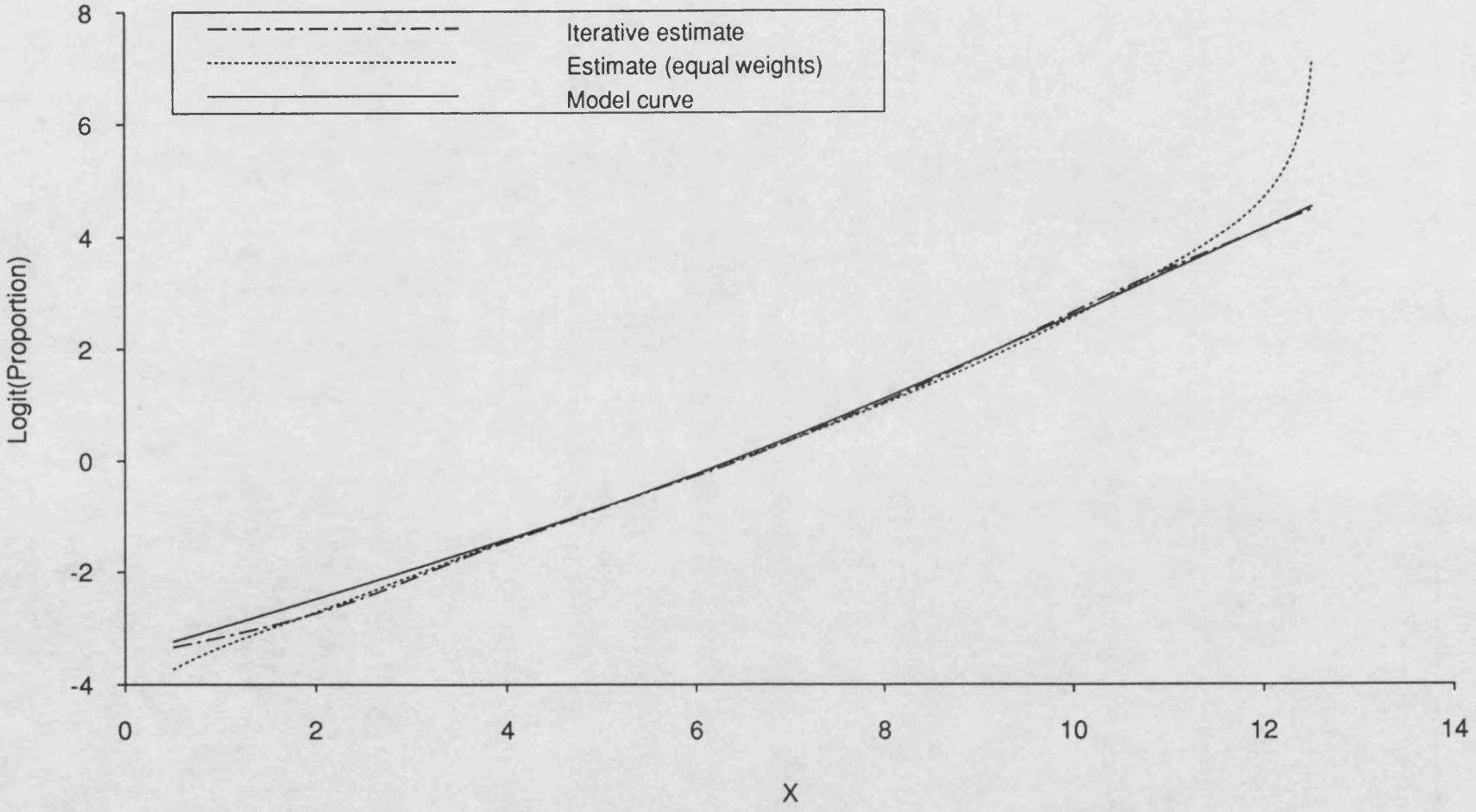


Fig. 6.2 shows, in addition, the spline estimate $\tilde{f}_{EW}$, (called the *equal weights* estimate) obtained by minimising

$$\sum_{i=1}^{60} w_i (g(x_i) - \bar{Y}_i)^2 + \alpha \int_a^b [g''(x)]^2 dx$$

subject to $g \in W_2$ and $\alpha > 0$. Here the weight attached to knot i (i=1,2,....60) is four and $\bar{Y}_i$ is the mean of the observations Y_{ik} at knot i. It will be seen that the resulting spline estimate $\tilde{f}_{EW}$ does not approximate the model curve at all well for $X < 4$ and $X > 11$ approximately. The corresponding graphs for the Gompertz calibration data set are given in Figs. 6.3 and 6.4. These figures show that the iterative spline estimate, $\hat{f}$, is an excellent estimate as it is very close to the model curve given by equation (6.2.19). Again Fig. 6.4 shows $\tilde{f}_{EW}$ to be a poor estimate at the ends of the calibration range (i.e. $X < 3$ and $X > 15$ approximately).

The conclusion of Sections 6.2.1 and 6.2.2 was that the posterior distribution of f approximates to a multivariate normal distribution which is the posterior distribution for the equivalent weighted least squares problem where pseudo-weights w_i^* are attached to pseudo-observations (x_i, Y_i^*) i=1,2,...60. Suppose m^* observations Y_j' are observed at the prediction stage of the calibration process corresponding to an *unknown* value of X, ξ. Let $Y_j^{*\prime} = \ln \dfrac{Y_j'}{(1 - Y_j')}$ and suppose $E(Y_j^{*\prime}) = \eta$ j=1,2,...m^*. It will be recalled from Section 3.2.2, Chapter 3, that the three steps in constructing the posterior distribution of ξ are as follows;

(i) Generate posterior realisations of f. Suppose such a realisation is denoted as $\tilde{f}_s$.

(ii) Generate posterior realisations of η. Suppose such a realisation is denoted as η_v.

(iii) Calculate $f_s^{-1}(\eta_v)$ for as many values of s and v as are required.

With particular reference to (i), the posterior distribution of f is a truncated multivariate normal distribution given by $\mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$ where

$$S = (\alpha\Omega + B^{\mathrm{T}} W^* B)$$

$$\hat{\gamma} = S^{-1} B^{\mathrm{T}} W^* \mathbf{Y}^*$$

Here $\mathbf{Y}^{*\mathrm{T}}$ is defined as $(\bar{Y}_1^* \bar{Y}_2^* \bar{Y}_3^* ... \bar{Y}_{60}^*)$ where $\bar{Y}_i^*$ is the mean of the pseudo-observations Y_i^* at knot i. Also W^* is a 60×60 diagonal matrix with entries w_i^* where w_i^* is the pseudo-weight attached to knot i.

Fig.6.3: Binomial proportions (Gompertz data set)

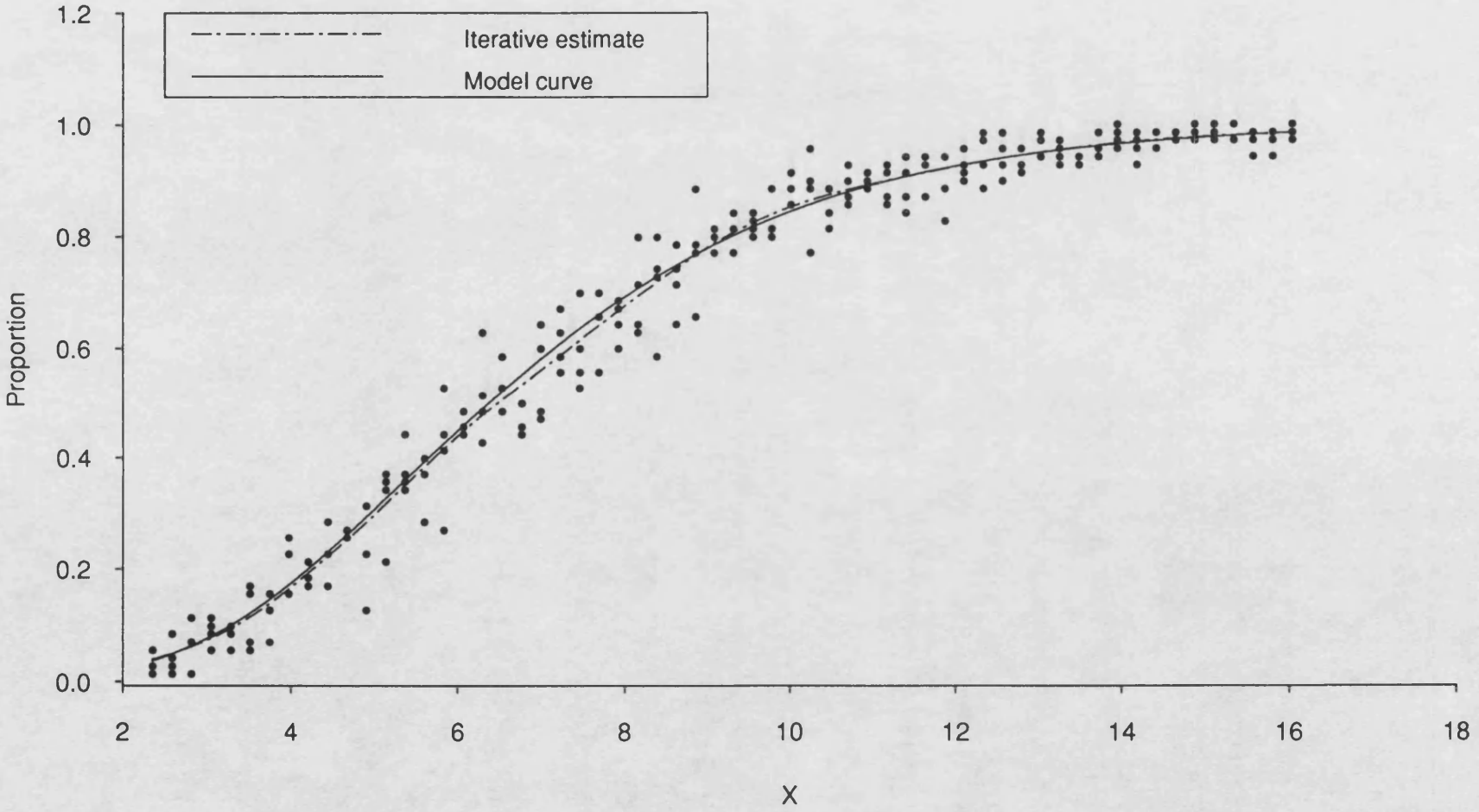
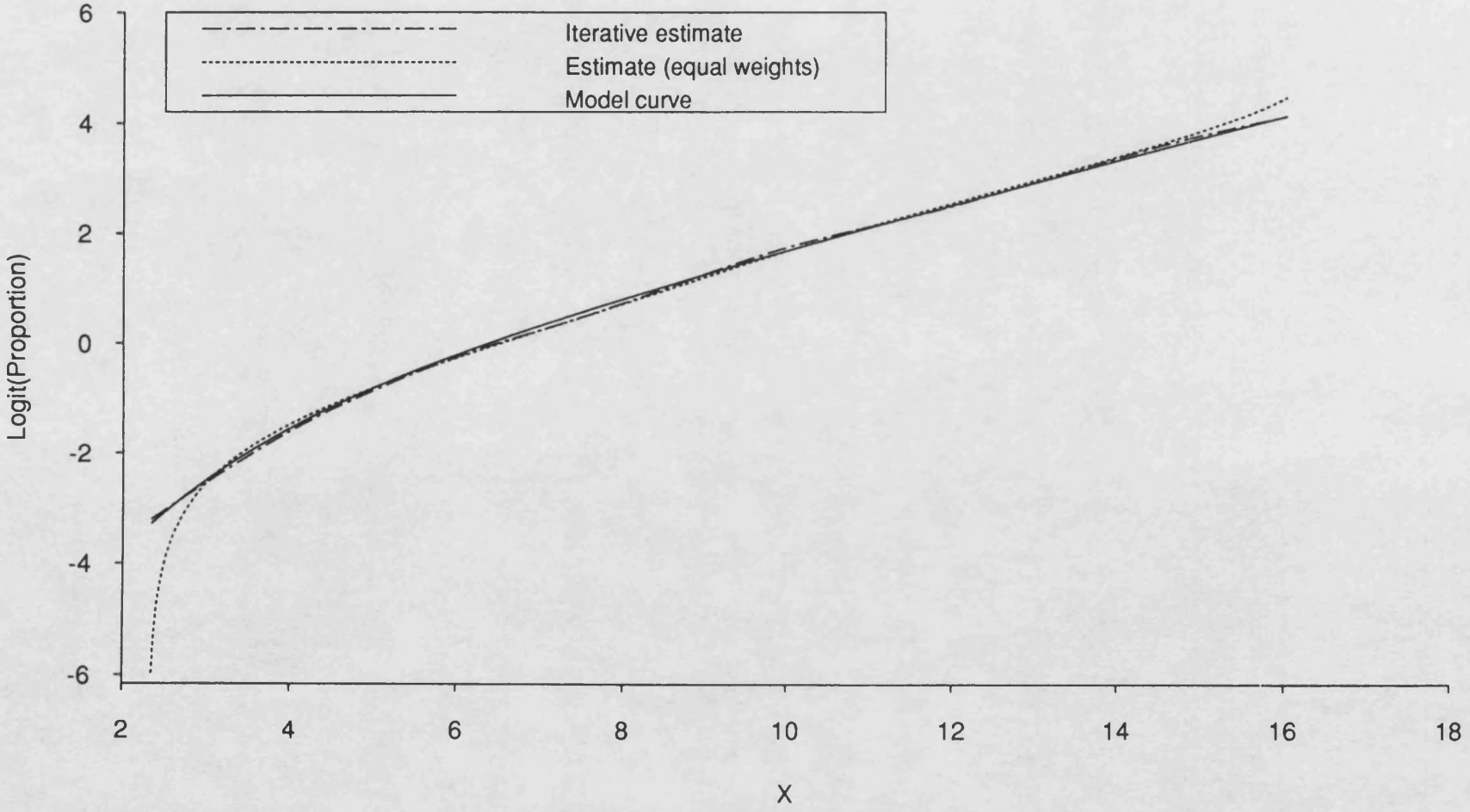


Fig.6.4: Transformed Binomial proportions (Gompertz data set)



With particular reference to (ii), the posterior distribution of η is given by

$$p(\eta|\mathbf{Y}^{*\prime}) = \left[\frac{2\pi\sigma^2}{m^*}\right]^{-\frac{1}{2}} \exp\left[\frac{-m^*(\eta-\bar{Y}^{*\prime})^2}{2\sigma^2}\right] \qquad \tilde{f}_s(x_1) \leq \eta \leq \tilde{f}_s(x_n)$$

$$= 0 \qquad \text{otherwise}$$

where $\bar{Y}^{*\prime}$ is the mean of the m^* observations $Y_j^{*\prime}$.

With particular reference to step (iii) above, suppose $s=1,2,\ldots,n$ and $v = 1,2,\ldots,m$ giving mn posterior realisations of ξ. It may recalled that in Chapter 3, Section 3.4, an equation for calculating the *optimal* m was given (equation 3.4.8). This was used for both data sets and n was calculated so that at least 10,000 posterior realisations of ξ were generated. For example, the optimal m for the first chosen value of X with the Richards data, was 14 so n was taken to be 715 giving a total of 100010 posterior realisations of ξ (10000/14 = 714.28).

To assess the performance of the method at the prediction stage of the calibration process, four observations Y_j' were simulated at each of twenty newly chosen values of X covering the calibration range. These were then transformed to the logit scale by taking $\ln \dfrac{Y_j'}{(1-Y_j')} = Y_j^{*\prime}$ $j=1,2,3,4$. Point estimates of ξ were obtained by using the posterior median and symmetrical interval estimates for ξ were constructed from the posterior distributions of ξ. An interval estimate was termed *successful* if the true value, ξ, lay inside it. Table 6.1 gives the mean absolute errors, the maximum width, the minimum width, the mean width and the proportion of successful estimates for the Gompertz and Richards data sets. Fig. 6.5 shows the point estimates, $\hat{\xi}$, plotted against the true value, ξ, the 45° line representing zero error. It will be noticed from Fig. 6.5 that errors are small.

6.3. A NON-PARAMETRIC APPROACH TO NON-BINOMIAL PROPORTIONS

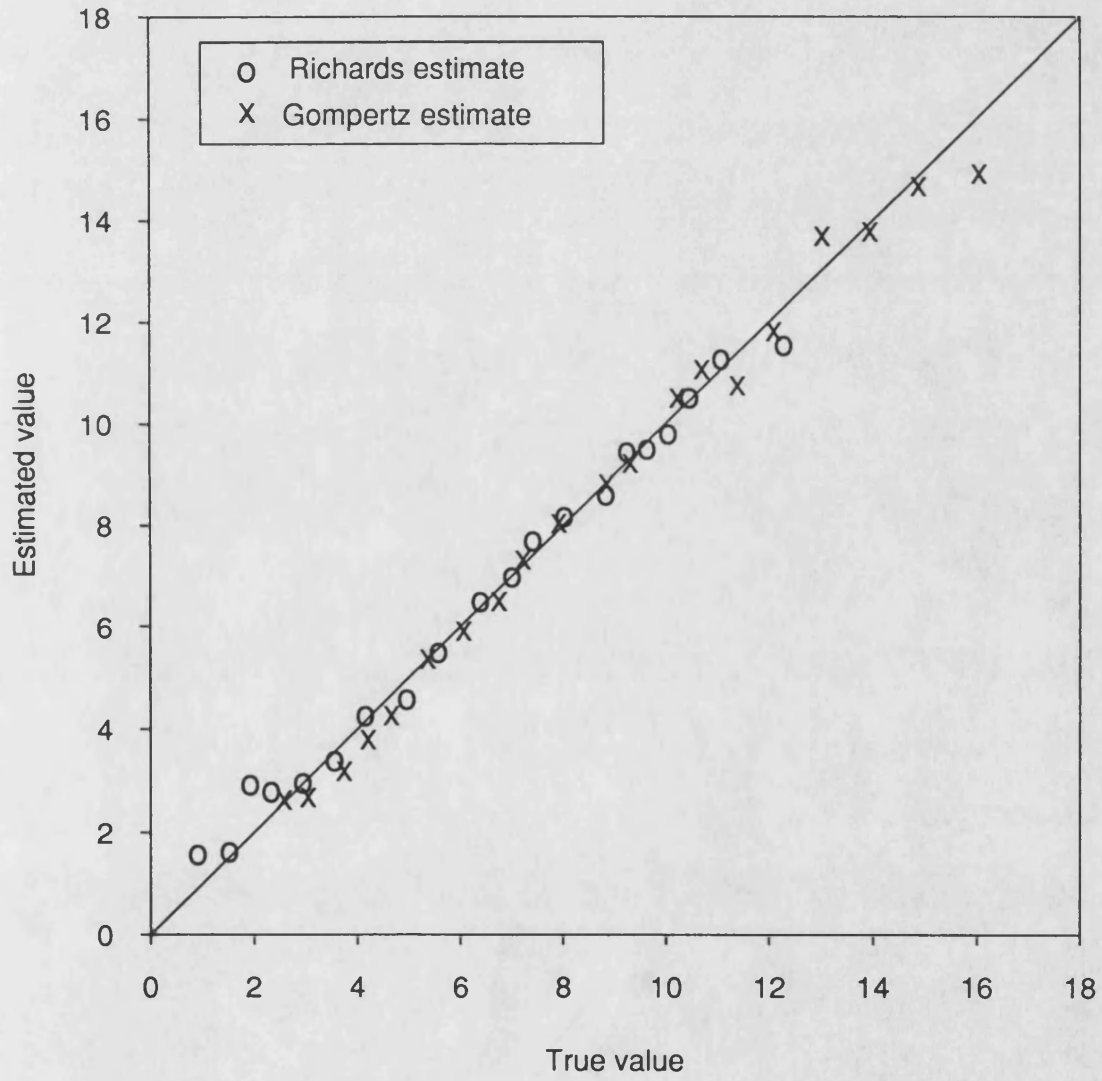
This section is concerned with an approach to proportion data where the proportion does not arise from a Binomial experiment. By way of illustration, this approach is applied to a teeth data set. Full details of the data set are given in Appendix 1. Measurements of the ratio of transparent root dentine to total root dentine were made on the upper canine teeth of seventeen patients. Three patients contributed two teeth, a left tooth *and* and right tooth, whereas the remaining patients contributed only one tooth, a left tooth *or* a right tooth. Let the subscript j indicate the number of teeth from each patient, then $j=1$ or 2. Three independent replications were made on each of the teeth; let the subscript k denote the kth replication, ($k=1,2,3$). The calibration data set therefore consisted

Table 6.1 Point and interval estimates for ξ (90% probability level)

	Data set	
	Gompertz	Richards
Mean absolute error	0.30	0.29
Maximum width of intervals	4.14	2.76
Minimum width of intervals	1.01	1.85
Mean width of intervals	2.97	2.38
No. of successful estimates	19	20
Total no. of estimates	20	20
Calibration range	2.35 - 16.05	0.5 - 12.5

Fig.6.5: Comparison of point estimates

(Richards and Gompertz data sets)



of 51 observations (x_i, R_{ijk}) where R_{ijk} represents the ratio of transparent dentine to total dentine on the jth tooth (kth replication) belonging to the ith patient. Since R_{ijk} is a ratio $0 < R_{ijk} < 1$ and so the question arises as to what distribution do we assume for R_{ijk}. Probably a suitable distribution would be a Beta distribution defined by

$$f(r) = \frac{\Gamma(\upsilon+\tau)}{\Gamma(\upsilon)\Gamma(\tau)} \; r^{\upsilon-1}(1-r)^{\tau-1} \quad \begin{matrix} 0 < r < 1 \\ \upsilon,\tau > 0 \end{matrix} , \qquad (6.3.1)$$

where r is the ratio of transparent root dentine to total root dentine. The problem with using R_{ijk} and assuming R_{ijk} is a Beta random variable with parameters υ and τ is that in so doing, it becomes impossible to apply the non-parametric methods of Chapters 3, 4 and 5 since the methods require that the regressor variable is a normal random variable and as such can take any value between $-\infty$ and $+\infty$, (see model 3.2.1). One is therefore forced into considering whether a transformation of R_{ijk} could overcome the two difficulties of

(a) R_{ijk} only taking values between 0 and 1,

(b) R_{ijk} having a Beta distribution, not a normal distribution.

Around the turn of the century, attempts were made to construct systems of distributions called *systems of frequency curves*, capable of representing a wider variety of observed frequency distributions than those for which a normal curve would suffice. The most well-known system is probably the Pearson system of frequency curves. In fact, the Beta distribution defined by equation (6.3.1) is a Pearson Type I distribution. A disadvantage of many of these systems of frequency curves is that many of the frequency curves are not well-approximated by the normal curve so it becomes difficult to apply normal theory in such cases.

However, Johnson (1949), put forward a system of frequency curves, derived by the method of translation, which are well approximated by the normal curve. The method of translation (Edgeworth, 1898) involves transforming variables such that the transformed variables may be considered to have a normal distribution. Johnson considered transforming from the variable Y to the variable Z where Z is a unit normal random variable and $Z = \gamma + \delta g(Y)$ where $g(Y)$ is a monotonic function of Y, whose range of values (corresponding to the actual range of possible values of Y) is $-\infty$ to $+\infty$. One of the transformations studied by Johnson was $g(Y) = \ln \frac{Y}{(1-Y)}$ $(0 < Y < 1)$. Johnson investigated the relation between the parameters γ and δ and the shape of the distribution of Y. Then he applied this transformation to variables with distributions of the Pearson system and commented on whether this resulted in a transformed variable following

approximately the normal law. Of particular interest is Johnson's application of the transformation $Z = \gamma + \delta \ln \frac{Y}{(1-Y)}$ to variables following a Pearson Type I distribution.

Let Y be a random variable with a Beta distribution with parameters υ and τ. Its density function is given by equation (6.3.1). Let $\sqrt{\beta_1}$ denote the coefficient of skewness of a distribution and β_2 denote the coefficient of kurtosis of a distribution. Johnson compared $\beta_1(Y)$, $\beta_2(Y)$ and $\beta_1(Z)$, $\beta_2(Z)$ with the values for a normal distribution, namely 0 and 3 respectively. Johnson obtained exact expressions for $\beta_1(Z)$, $\beta_2(Z)$ and stated that for υ and τ sufficiently large

$$\beta_1(Z) \approx \upsilon^{-1} + \tau^{-1} - 4(\upsilon+\tau)^{-1}$$

$$\beta_2(Z) \approx 3 + 2\upsilon^{-1} + 2\tau^{-1} - 6(\upsilon+\tau)^{-1} \qquad (6.3.2)$$

Equations (6.3.2) were compared with 0 and 3 and also with the approximate expressions for $\beta_1(Y)$, $\beta_2(Y)$ given by

$$\beta_1(Y) \approx 4\upsilon^{-1} + 4\tau^{-1} - 16(\upsilon+\tau)^{-1}$$

$$\beta_2(Y) \approx 3 + 6\upsilon^{-1} + 6\tau^{-1} - 30(\upsilon+\tau)^{-1}$$

These comparisons showed that $\ln \frac{Y}{(1-Y)}$ followed more closely a normal law than did Y. Johnson also compared $\beta_1(Z)$, $\beta_2(Z)$ defined by equations (6.3.2) with the approximate expressions for $\beta_1(Z)$, $\beta_2(Z)$ when the transformation $Z = \gamma + \delta \ln Y$ is applied to the same Type I variable Y. These are given by

$$\beta_1(Z) \approx \upsilon^{-1} + 4\tau^{-1} - (\upsilon+\tau)^{-1}$$

$$\beta_2(Z) \approx 3 + 2\upsilon^{-1} + 6\tau^{-1} - 2(\upsilon+\tau)^{-1}$$

These comparisons showed that the transformation $Z_B = \gamma + \delta \ln(Y/(1-Y))$ generally produced a closer approach to normality than the transformation $Z_L = \gamma + \delta \ln Y$.

Returning to the teeth data it was decided, in view of the findings of Johnson's paper, to transform R_{ijk} by taking $Q_{ijk} = \ln(R_{ijk}/(1-R_{ijk}))$. This transformation had the effect of also stabilising the variance $Var(R_{ijk}|X=x)$. It can be assumed that Q_{ijk} has an approximately normal distribution and so now the non-parametric methods described in Chapters 3, 4 and 5 can be applied to this transformed data set, to produce the posterior distribution of ξ, where ξ is the age of a *new* patient. By way of illustration, the method developed in Chapter 3 is applied to the transformed data. Suppose there are m patients with ages $x_1, x_2, \ldots, x_m$ in the calibration data set, each contributing one or two teeth. The

mathematical model for the calibration experiment is the following nested model

$$Q_{ijk} = f(x_i) + \tau_i + \eta_{ij} + \varepsilon_{ijk} \qquad \begin{matrix} i=1,2,\ldots,m \\ j=1 \text{ or } 2 \\ k=1,2,3 \end{matrix} \tag{6.3.3}$$

where Q_{ijk} = transformed ratio of transparent root dentine to total root dentine for the jth tooth of the ith patient (kth replication), τ_i is the effect due to the ith patient, η_{ij} is the left/right tooth effect, and f is a monotonic natural cubic smoothing spline with knots $\{x_i\}$ $i=1,2,\ldots,m$. It is assumed that the τ_i are i.i.d. $N(0,\sigma_\tau^2)$, the η_{ij} are i.i.d. $N(0,\sigma_\eta^2)$ and the ε_{ijk} are i.i.d. $N(0,\sigma_\varepsilon^2)$. Also it is assumed that the τ_i, η_{ij} and ε_{ijk} are independent of each other.

Using the model defined in equation (6.3.3),

$$\text{cov}\,(Q_{ijk}, Q_{ij'k}) = \begin{cases} \sigma_\tau^2 & j \neq j' \\ \sigma_\tau^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 & j=j' \end{cases}$$

and $\text{cov}\,(Q_{ijk}, Q_{i'j'k'}) = 0$ if $i \neq i'$ and $\text{cov}\,(Q_{ijk}, Q_{i'j'k'}) = 0$ if $k \neq k'$. Obviously measurements on the right and left teeth from the *same* patient are correlated. Suppose this is measured by the correlation coefficient ρ, then

$$\rho = \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma_\eta^2+\sigma_\varepsilon^2} = \frac{\sigma_\tau^2}{\sigma^2}$$

where $\text{Var}(Q_{ijk}) = \sigma_\tau^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 = \sigma^2$. Let $\varepsilon_{ijk}^* = \tau_i + \eta_{ij} + \varepsilon_{ijk}$ then model (6.3.3) can be written as follows:

$$Q_{ijk} = f(x_i) + \varepsilon_{ijk}^* \qquad \begin{matrix} i=1,2,\ldots,m \\ j=1 \text{ or } 2 \\ k=1,2,3 \end{matrix}$$

where the ε_{ijk}^* are normally distributed random variables. It may be recalled from Section 3.2.2, Chapter 3, that $f(x) = \sum_{i=1}^{n} \gamma_i \beta_i(x)$ (equation 3.1.12) where $\{\beta_1,\beta_2,\ldots,\beta_n\}$ are the B-spline basis. The posterior distribution of γ is a multivariate normal truncated to exclude those values of γ which produce non-monotonic splines. Using the notation of Section 3.2.2, $\gamma \sim \mathcal{N}_{MON}(\hat{\gamma}, S^{-1})$ where

$$S^{-1} = \sigma^2(B^{\mathrm{T}}WB + \alpha\Omega)^{-1}$$

$$\hat{\gamma} = (B^{\mathrm{T}}WB + \alpha\Omega)^{-1} B^{\mathrm{T}}W\mathbf{Y}.$$

Suppose for simplicity that all the m ages $x_1,x_2,\ldots,x_m$ are distinct, then W is an $m \times m$ diagonal matrix with entries w_i. Now for the case of $j = 2$

$$\text{Var}(\bar{Q}_{i..}) = \text{Var}\left[\sum_{j=1}^{2}\sum_{k=1}^{3}\frac{Q_{ijk}}{6}\right]$$

$$= \frac{1}{6}\sigma^2(1+3\rho)$$

where $\bar{Q}_{i..}$ is the mean of the observations at knot i (i=1,2,...,m). So the weight attached to knot i when j=2 is $6/(1+3\rho)$. For the case of j=1, $\text{Var}(\bar{Q}_{i..}) = \sigma^2/3$ so the weight attached to knot i when j=1 is three. Hence the diagonal matrix W has entries w_i where

$w_i = 6/(1+3\rho)$ if patient i contributes two teeth

$= 3$ if patient i contributes one tooth.

It was necessary to choose the smoothing parameter α and estimate ρ and σ^2. This was achieved by using the BATHSPLINE computer package (Silverman and Watters, 1984). The correlation coefficient ρ was estimated using an iterative procedure similar to that described in Section 3.3.2, Chapter 3.

Let $d_{i.} = \bar{Q}_{i1.} - \bar{Q}_{i2.}$. From model (6.3.3), $d_{i.} = (\eta_{i1}-\eta_{i2}) + (\bar{\varepsilon}_{i1.}-\bar{\varepsilon}_{i2.})$ so

$$d_{i.} \sim N(0,\ 2(\sigma_\eta^2+\sigma_\varepsilon^2/3)) = N(0,\sigma_d^2)$$

An estimate of σ_ε^2 is given by (Scheffé, 1959)

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{2}\sum_{k=1}^{3}(Q_{ijk}-\bar{Q}_{ij.})^2}{\sum_i\sum_j(2)} \qquad (6.3.4)$$

and so an estimate of σ_η^2 is given by

$$\hat{\sigma}_\eta^2 = \frac{\hat{\sigma}_d^2}{2} - \frac{\hat{\sigma}_\varepsilon^2}{3} = \frac{1}{2(m_2-1)}\left\{\sum_{i=1}^{m_2} d_{i.}^2 - \frac{(\Sigma d_{i.})^2}{m_2}\right\} - \frac{\hat{\sigma}_\varepsilon^2}{3} \qquad (6.3.5)$$

where m_2 is the number of patients contributing two teeth. The stages of the iterative process for estimating ρ are given by

(i) Calculate an estimate of σ^2 using the BATHSPLINE package with weight w_i attached to knot i.

(ii) Calculate an estimate of σ_ε^2 and σ_η^2 using equations (6.3.4) and (6.3.5) respectively.

(iii) Obtain an estimate of ρ using

$$\hat{\rho} = \frac{\hat{\sigma}^2 - \hat{\sigma}_\varepsilon^2 - \hat{\sigma}_\eta^2}{\hat{\sigma}^2}$$

(iv) Calculate the weights $w_i = \dfrac{6}{(1+3\hat{\rho})}$

Steps (i) - (iv) were repeated until successive estimates of ρ agreed to three decimal places. For the first iteration *only*, the w_i were all taken to be one.

Fig. 6.6 shows the calibration data set together with the spline smoother $\hat{f}_\alpha$. The model for the prediction stage is given by

$$Q_k' = f(\xi) + \varepsilon_k' \qquad k=1,2,3.$$

Here Q_k' is the transformed ratio of transparent root dentine to total root dentine on the tooth of a *new* patient of age ξ years. It is assumed that the ε_k' are i.i.d. $N(0,\sigma^2)$ random variables. The posterior distribution of ξ was obtained using the method described in Chapter 3. The posterior median was used as a point estimator of ξ and symmetrical interval estimates were constructed from the posterior distribution for ξ assuming an 80% probability level. Again interval estimates were termed *successful* if the true value, ξ, lay within the interval. Table 6.2 gives the mean absolute error and details of the interval estimates. The results are very satisfactory from an odontological point of view.

6.4 CONCLUDING REMARKS

This chapter has considered two non-parametric approaches to proportions. Where the proportion arises from a Binomial model, a non-parametric approach to logistic regression was derived. The optimal solution is $\hat{f}$, a natural cubic spline with knots at the data points $\{x_i\}$ $i=1,2,\ldots,n$ and $\hat{f}$ is calculated by using iterative formulae similar to those used with generalised linear models. Having obtained the iterative spline estimate, $\hat{f}$, the methods described in Chapters 3, 4 and 5 can be used to generate the posterior distribution of ξ. The approach was tested out on two simulated data sets and produced good estimates, $\hat{f}$, and good point and interval estimates for ξ. The approach was easy to implement and very efficient in terms of computer time. There seemed to be no obvious disadvantages.

The second approach to proportions is also very easy to implement because it only involves transforming the data, using a logit transformation, and then applying the methods described in Chapters 3, 4 and 5 to produce the posterior distribution of ξ. The approach was applied to a teeth data set, consisting of proportions of transparent root dentine and ages of patients and seemed to produce reasonable results.

Fig.6.6: Upper canine teeth

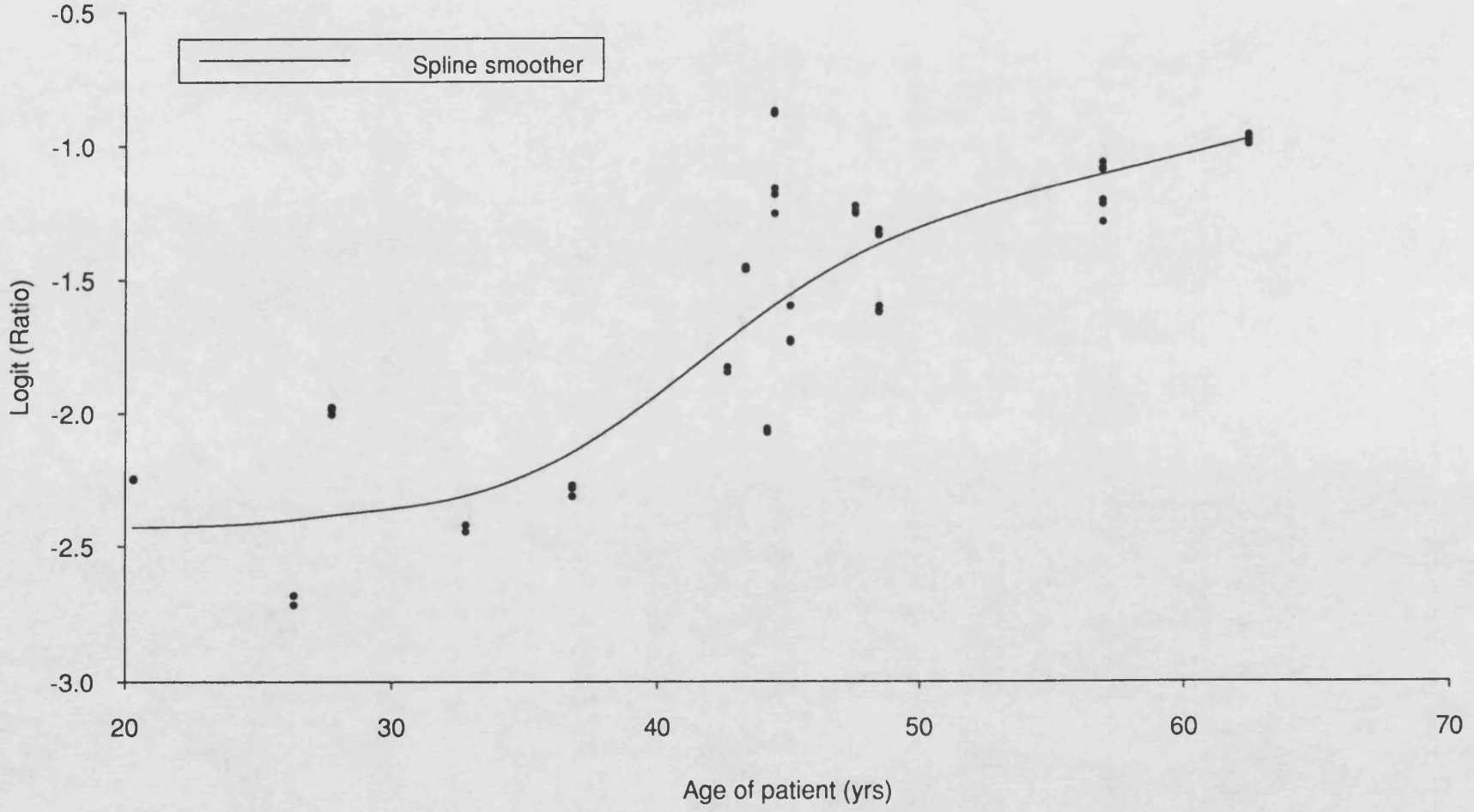


Table 6.2 Mean absolute error and interval estimates for ξ for the upper canine teeth (80% probability level)

Mean absolute error (years)	5.9
Maximum width of interval estimates (years)	11.26
Minimum width of interval estimates (years)	5.32
Mean width of interval estimates (years)	8.67
No. of successful estimates	4
Total number of estimates	7

7. COMPARISONS AND CONCLUSIONS

7.1 COMPARISONS OF THE NON-PARAMETRIC METHODS

This section compares the three approaches detailed in Chapters 3, 4 and 5. Firstly the theories of each approach are compared and then the results obtained from applying each of the methods to all of the data sets.

7.1.1. Comparison of methodologies

It will be recalled that with the approaches of Chapters 3 and 4, ξ was viewed as a non-linear functional of the true calibration curve, f, and simulation was used to obtain the posterior distribution of ξ whereas with the approach developed in Chapter 5, predictive densities were used. Let us consider the models for the approaches of Chapters 3 and 4. The models for the calibration experiment and prediction stage are given by

$$\textit{Calibration experiment:} \qquad Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,n \qquad (7.1.1)$$

$$\textit{Prediction stage:} \qquad Y_j' = f(\xi) + \varepsilon_j'$$

$$= \eta + \varepsilon_j' \qquad j=1,2,\ldots m \qquad (7.1.2)$$

where f is a monotonic natural cubic spline with knots $\{x_i\}$ $i=1,2,\ldots,n$, the ε_i are i.i.d. normal random variables with mean zero and variance σ^2/w_i and the ε_j' are i.i.d. normal random variables with mean zero and variance σ^2. It is also assumed that the ε_i are independent of the ε_j'. Suppose the approach developed in Chapter 3 is denoted by S (simulation) whilst the approach developed in Chapter 4 is denoted by SAS (switching algorithm and simulation). The three basic steps to obtaining the posterior distribution of ξ using method S are as follows:

(i) Simulate from the posterior distribution of f which is a truncated multivariate normal distribution. Suppose such a posterior realisation of f is called $\tilde{f}_s$.

(ii) Simulate from the posterior distribution of η which is a truncated normal distribution. Suppose such a posterior realisation of η is called η_v.

(iii) Calculate $\tilde{f}_s^{-1}(\eta_v)$ for as many values of v and s as are required.

With method SAS, the first step is the same as that of method S but then a switching algorithm is used to obtain the posterior distribution of ξ. Suppose ξ has prior density $\pi(\xi)$ and that this density is unimodal within $[x_1, x_n]$. Let $\pi_{MAX} = \underset{x_1 \le \xi \le x_n}{Max}\, \pi(\xi)$. Theorem 4.1 shows that if $p(\eta|\mathbf{Y}')$ is the posterior density

of η corresponding to the prior density $\pi(\xi)$ for ξ, then realisations of $p(\eta|\mathbf{Y}')$ can be obtained by

(i) simulating η^* from the posterior density $p_u(\eta|\mathbf{Y}')$, which is the posterior density of η corresponding to a uniform prior density for ξ

(ii) accepting η^* as a realisation of a random variable with p.d.f. $p(\eta|\mathbf{Y}')$ iff

$$U\pi_{MAX} \leq \pi(\tilde{f}_s^{-1}(\eta^*))$$

where $\tilde{f}_s$ is a posterior realisation of f and U is an independent $U(0,1)$ random variable.

The posterior distribution of ξ therefore consists of the values ξ^* where $\xi^* = \tilde{f}_s^{-1}(\eta^*)$. A switching algorithm based on a modification of the switching algorithm of Atkinson and Whittaker (1976) was used to simulate η^* from the posterior density $p_u(\eta|\mathbf{Y}')$.

If one compares methods S and SAS, they both consider ξ to be a non-linear functional of f and they both use simulation to obtain the posterior distribution of ξ but method SAS enables prior information on ξ, in the form of a prior density for ξ, to be easily incorporated. With method S, one assumes a prior for η which is given by

$$\begin{aligned} p(\eta) &= c \quad \tilde{f}_s(x_1) \leq \eta \leq \tilde{f}_s(x_n) \\ &= 0 \quad \text{otherwise} \end{aligned}$$

where c is a constant. This will imply a prior for ξ since $\xi = f^{-1}(\eta)$.

The approach considered in Chapter 5 is completely different from those of Chapters 3 and 4 because it does not use simulation but is based on predictive densities. Also there is no conditioning on monotonicity. The models for the calibration experiment and predictive stage are given respectively by

$$\begin{aligned} \textit{Calibration experiment:} \quad & Y_i = f(x_i) + \varepsilon_i \qquad i=1,2,\ldots,n \\ \textit{Prediction stage:} \quad & Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,\ldots m \end{aligned}$$

where it is assumed that f is a natural cubic smoothing spline with knots $\{x_i\}$ $i=1,2,\ldots,n$. The assumptions about ε_i and ε_j' are the same as those for models (7.1.1) and (7.1.2).

The residual sum of squares (RSS) for the calibration experiment, in the case of the w_i being all one, is given by equations (5.1.2) and (5.1.4), namely

$$\begin{aligned} \text{RSS} &= (\mathbf{Y} - \hat{\mathbf{f}})^{\mathrm{T}}(\mathbf{Y} - \hat{\mathbf{f}}) \\ &= \mathbf{f}^{\mathrm{T}}D\mathbf{f} + 2\mathbf{f}^{\mathrm{T}}D\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^{\mathrm{T}}D\boldsymbol{\varepsilon} \end{aligned}$$

where $D = (I-A)^2$, A is the *hat matrix* and $\mathbf{f}^T = (f(x_1)\, f(x_n),\ldots,f(x_n))$. It was argued that provided n was reasonably large, the distribution of RSS could be approximated by the distribution of $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$ which is a quadratic form in normal random variables. The most successful approximation considered was the Solomon-Stephens approximation which involved approximating the distribution of $\boldsymbol{\varepsilon}^T D\boldsymbol{\varepsilon}$ by $c(\chi^2_v)^d$ with the parameters c, d and v being chosen to equate the first three moments of the two distributions. This approximation to the distribution of RSS enabled the predictive density function $L(\xi) = p(\bar{Y}' \,|\, \xi, \mathbf{z}, V_f)$ to be evaluated, where $\bar{Y}'$ is the mean of the m observations $Y_1', Y_2', \ldots\ldots, Y_m'$, $\mathbf{z}$ is the calibration data set $\{(x_i, Y_i)\ i=1,2,\ldots,N\}$ and $V_f = \sum_{j=1}^{m} (Y' - \bar{Y}')^2$. The evaluation of $L(\xi)$ involved numerical integration. The posterior density of ξ, $\pi(\xi \,|\, \mathbf{z}, \mathbf{Y}')$, is given by equation (5.2.6), namely

$$\pi(\xi \,|\, \mathbf{z}, \mathbf{Y}') \propto \begin{cases} \pi(\xi \,|\, \mathbf{x})\, L(\xi) & \text{if calibration experiment is random} \\ \pi(\xi)\, L(\xi) & \text{if calibration experiment is controlled} \end{cases} \tag{7.1.3}$$

Where f was assumed to be an interpolating rather than a smoothing spline, an approximation to the distribution of RSS was not necessary and $L(\xi)$ was shown to be a non-central Student density function (see Section 7.1.2). Suppose this predictive density approach is denoted by PD.

Methods SAS and PD are similar in the sense that both require the specification of a prior density for ξ and hence directly utilise prior information on ξ, in contrast to method S which is not able to do this. All the three methods were easy to implement and were efficient in terms of computer time. There appeared to be no obvious disadvantages with any of the methods. Let us now compare the results obtained when the three methods are applied to the antibiotic assay data and the five simulated data sets Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic.

7.1.2 Comparison of methods S, SAS and PD applied to the antibiotic assay data

In bioassays, enzyme assays or radio immunoassays, the concentration response relationships are mostly non-linear. It is frequently argued with such assays, that prediction samples can be diluted so one only need consider the *almost linear* part of the response curve. Racine-Poon (1988) disputes this as she states that very often because of physical restrictions or for biological reasons, prediction samples cannot be diluted so one must consider a non-linear response

curve (f). Frequently, scientists wish to estimate very low concentrations of a substance and this often corresponds to the curvilinear part of the response curve (see Fig. 3.20 for an example of this), making standard linear calibration methods inappropriate. All the non-parametric methods S, SAS and PD were applied to antibiotic assay data with a fair degree of success.

Using the predictive density approach of Chapter 5 on this data set did not require an approximation to the distribution of RSS. The predictive density function, $L(\xi)$, is given by equation (5.6.4), namely

$$L(\xi) = St\left[\upsilon,\ \beta^{T}\hat{\gamma},\ \frac{VC_{\xi}}{\upsilon}\right]$$

i.e. $L(\xi)$ is the density function (non-normalised) of a non-central Student t-distribution based on υ degrees of freedom, centred at $\beta^{T}\hat{\gamma}$ and with scale factor VC_{ξ}/υ. When applying methods SAS and PD, a uniform prior density for ξ was assumed. Table 7.1 gives details of the point and interval estimates for ξ obtained using the three methods. It may be recalled that for methods S and SAS, the posterior median was used as a point estimator of ξ and symmetrical interval estimates were calculated for ξ whereas for method PD, it was recommended that the posterior mode was used as a point estimator of ξ and highest posterior density intervals were used as interval estimates for ξ.

Table 7.1. indicates that all three methods produce good results, particularly when one considers that these are based on only *one* observation being made at the prediction stage. If one compares the methods, the lowest mean absolute error is that obtained with method PD. With respect to interval estimation, methods SAS and PD give very similar results. With a probability level of 90%, one would expect *on average* about 54 successful estimates. The number of successful estimates for method S is just acceptable but it is substantially lower than that obtained with methods SAS and PD. It would seem therefore, that either method, SAS or PD, would give a quick and efficient method of estimation of concentrations. Method SAS should be used for the case of f being a smoothing spline and method PD used for the case of f being an interpolating spline.

7.1.3. Comparison of methods S, SAS and PD applied to the simulated data sets

All three methods S, SAS and PD were applied to the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets and the details of the resulting point estimates of ξ are given in Table 7.2. For methods SAS and PD it

Table 7.1 Comparison of all the methods (Antibiotic assay data)

	Method		
	S	SAS	PD
Mean absolute error	0.069	0.069	0.061
Mean width of interval estimates (90% probability level)	0.30	0.30	0.29
No. of successful estimates	52	57	57
Total no. of estimates	60	60	60
Calibration range	−0.22 to 1.00		

Table 7.2 Comparison of mean absolute errors when assuming a uniform prior for ξ

Data set	Method	Mean absolute error	Calibration range
Gompertz	S	0.39	
	SAS	0.35	0 – 13.43
	PD	0.38	
Weibull	S	0.72	
	SAS	0.71	10.0 – 79.52
	PD	0.69	
Preece-Baines	S	0.11	
	SAS	0.15	4 – 19.8
	PD	0.12	
Bleasdale-Nelder	S	3.15	
	SAS	3.18	20 – 178
	PD	3.20	
Asymptotic	S	1.04	
	SAS	1.10	1 – 21.74
	PD	1.11	

was assumed that ξ had a uniform prior density. Table 7.2 shows that all three methods produce good point estimates of ξ. A comparison of the mean absolute errors reveals that all three methods are equally good. Figs. 7.1 and 7.2 show the estimated value of ξ, $\hat{\xi}$, plotted against the true value, ξ, for the Bleasdale-Nelder and Asymptotic data sets respectively. The 45° line represents zero error. Fig. 7.1 shows that for all three methods errors are very small and the point estimates are virtually identical. Fig. 7.2 reveals a pleasing property of all three methods, namely that where the data are less informative (i.e. the spline is flatter) the errors are larger. For the Asymptotic data set this is for values of $\xi > 11$ approximately (see Fig. 3.11). Also Fig. 7.2 reveals that differences in the point estimates obtained using the three methods are most pronounced when the data are less informative. This is partly due to the fact that different point estimators are used for methods S, SAS and PD and differences between the posterior mode and posterior median are accentuated where the data are less informative because of the resulting shape of the posterior distribution of ξ.

Table 7.3 gives the details of the interval estimates for ξ obtained using all three methods and a uniform prior density for ξ. The table shows that all three methods give good interval estimates for ξ. However, the interval estimates obtained using methods S and SAS are more precise. This is almost certainly because of the approximation to the distribution of RSS with method PD. If one refers back to Table 7.1, one will see that where an approximation to the distribution of RSS was *not* necessary, because its exact distribution was known (see Section 5.6, Chapter 5), the interval estimates for all three methods were *equally* precise. If one were able to improve the approximation to the distribution of RSS, one should be able to reduce the difference in precision of estimates between methods S, SAS and method PD. This would be a good topic for future research. With reference to the proportion of successful estimates, Table 7.3 reveals that all three methods are equally good.

Let us now compare methods S and SAS because as was stated in Section 7.1.1, both methods are based on simulation and use the theory of estimation of non-linear functionals. It will be recalled that method S possesses many similarities with Hunter and Lamboy's (1981a) approach to linear calibration (compare Figs. 3.4 and 3.6). Hunter and Lamboy assumed a locally uniform prior for η which implied a prior for ξ which was vague (non-informative) and rather flat. Tables 7.4 and 7.5 give comparisons of methods S and SAS when assuming respectively a triangular prior density for ξ, defined by equation (4.6.1) and a truncated non-central Student prior density for ξ, defined by equation (4.6.3). Comparison of Tables 7.2, 7.3, 7.4 and 7.5 shows that method S is closest to

Fig.7.1: Comparison of point estimates
(Bleasdale-Nelder data set)

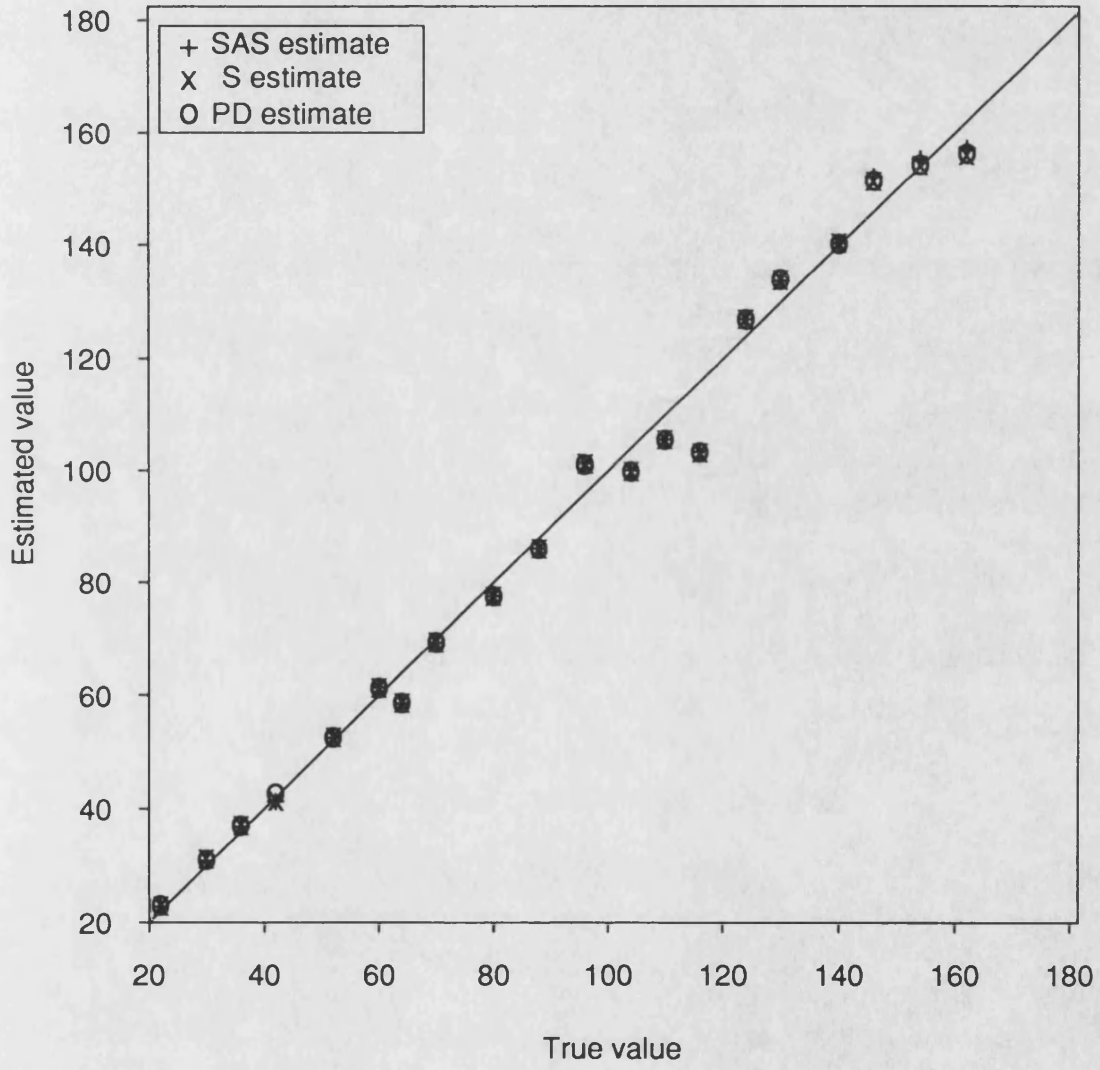


Fig.7.2: Comparison of point estimates
(Asymptotic data set)

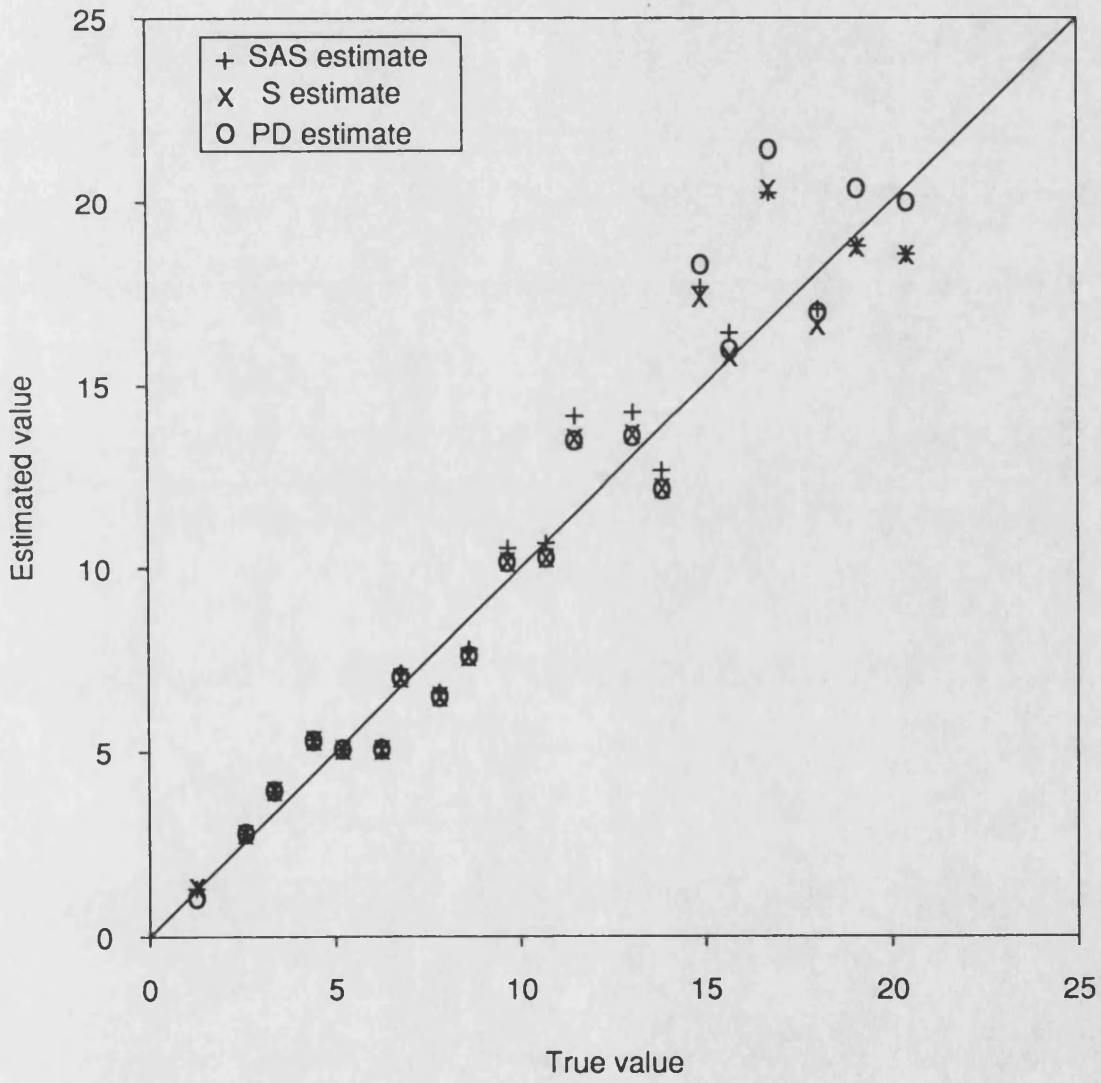


Table 7.3 Comparison of interval estimates when assuming a uniform prior for ξ (90% probability level)

Data set	Method	Mean interval width	Proportion of successful estimates
Gompertz	S	1.56	19/20
	SAS	1.57	19/20
	PD	1.77	19/20
Weibull	S	3.53	19/20
	SAS	3.55	19/20
	PD	4.42	19/20
Preece-Baines	S	0.72	19/20
	SAS	0.74	19/20
	PD	0.90	19/20
Bleasdale-Nelder	S	13.14	18/20
	SAS	13.25	18/20
	PD	16.14	18/20
Asymptotic	S	4.96	19/20
	SAS	4.95	19/20
	PD	5.59	19/20

Table 7.4 Comparison of methods S and SAS when assuming a triangular prior for ξ

Data set	Method	Mean absolute error	Mean width of interval estimates	Proportion of successful estimates
Gompertz	S	0.39	1.56	19/20
	SAS	0.40	1.44	18/20
Preece -Baines	S	0.11	0.72	19/20
	SAS	0.13	0.70	18/20
Bleasdale -Nelder	S	3.15	13.14	18/20
	SAS	3.12	12.70	17/20

Table 7.5 Comparison of methods S and SAS when assuming a non-central Student prior for ξ

Data set	Method	Mean absolute error	Mean width of interval estimates	Proportion of successful estimates
Weibull	S	0.72	3.53	19/20
	SAS	0.72	3.52	19/20
Asymptotic	S	1.04	4.96	19/20
	SAS	1.12	4.84	19/20

method SAS when one assumes a uniform prior for ξ with method SAS, i.e. method S is almost equivalent to method SAS if one assumes a uniform prior density for ξ on $[x_1, x_n]$ with method SAS. This seems reasonable in the light of what has been said above, concerning the implications for ξ of Hunter and Lamboy's choice of prior for η. Study of the results in Tables 7.2 and 7.3 would seem to suggest that if the true calibration curve, f, is assumed to be a monotonic natural smoothing spline and there is no prior information ξ or only vague prior information about ξ, method S should be used in preference to method SAS.

Finally in this sub-section, methods SAS and PD are compared. With both methods, prior information on ξ is utilised in the form of a prior density for ξ. Tables 7.6 and 7.7 show details of the estimates for ξ when using both methods, assuming firstly a triangular prior density for ξ given by equation (4.6.1) and secondly a truncated non-central Student prior density for ξ given by equation (4.6.3). These tables show that with respect to point estimation of ξ, method SAS is better than method PD, producing a smaller mean absolute error. Also, with respect to precision of interval estimates, method SAS is better. Although for four out of five of the data sets, the proportion of successful estimates is higher for method PD, this is almost certainly due to the fact that intervals are wider with method PD. Method SAS is to be preferred to method PD, particularly when one looks at the results given in Table 7.3, where a uniform prior density for ξ is assumed.

7.1.4 Comparison of methods S, SAS and PD where the number of knots is reduced

As already mentioned in Section 7.1.1, method PD involves an approximation to the distribution of the residual sum of squares (RSS) for the calibration experiment, if one assumes the true calibration curve f is a natural cubic smoothing spline. In the development of method PD, it was argued that for large n, the distribution of RSS could be approximated by the distribution of $\varepsilon^T D \varepsilon$, which is a quadratic form in normal random variables. Here n is the number of knots of the smoothing spline. With the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic data sets analysed in Chapters 3, 4 and 5, the number of knots was eighty and the total number of observations in each calibration data set (N) was 240. It was decided that it would be interesting to reduce the number of knots from eighty to forty and N from 240 to 40, to see how all the methods performed, but in particular how method PD performed. To this end, observations were simulated from the Gompertz and Weibull models (see Appendix 1 for further details). Suppose these data sets are denoted by Gp40 and

Table 7.6 Comparison of methods SAS and PD when assuming a triangular prior for ξ

Data set	Method	Mean absolute error	Mean width of interval estimates	Proportion of successful estimates
Gompertz	SAS	0.40	1.44	18/20
	PD	0.45	1.65	19/20
Preece -Baines	SAS	0.13	0.70	18/20
	PD	0.14	0.87	19/20
Bleasdale -Nelder	SAS	3.12	12.70	17/20
	PD	3.25	15.34	17/20

Table 7.7 Comparison of methods SAS and PD when assuming a non-central Student prior for ξ

Data set	Method	Mean absolute error	Mean width of interval estimates	Proportion of successful estimates
Weibull	SAS	0.72	3.52	19/20
	PD	0.74	4.52	20/20
Asymptotic	SAS	1.12	4.84	19/20
	PD	1.25	5.59	20/20

Wb40 respectively. To assess the performance of the three methods at the prediction stage of the calibration process, three observations Y_j' were simulated at each of twenty newly chosen values of X covering the calibration range, so that the model for the prediction stage is given by

$$Y_j' = f(\xi) + \varepsilon_j' \qquad j=1,2,3.$$

Both point and interval estimates for ξ were obtained using all three methods. A uniform prior density for ξ was assumed with methods SAS and PD. Table 7.8 gives details of the point estimates for all three methods. The mean absolute errors have been given for method PD when using both the posterior mode and posterior median as point estimators of ξ. These results are very pleasing because the mean absolute errors are still small despite the total number of observations in the calibration experiment (N) being reduced from 240 to 40. The mean absolute errors are only slightly larger than those obtained when N = 240 (see Table 7.2). Fig. 7.3 shows the estimated value of ξ, $\hat{\xi}$, plotted against the true value, ξ, for all the methods applied to the Gp40 data set. This plot shows the errors to be small except where the data are less informative concerning the value of ξ, which is for values of ξ approximately < 3 and > 11 (c.f. Fig. 3.7 where N = 240). This is a good property for any method to have. Table 7.9 gives details of the interval estimates for ξ for all the methods applied to the two data sets. In the case of method PD, values are given for both highest posterior density interval estimates (HP) and symmetrical interval estimates (SY). The interval estimates are good being only a little less precise than those obtained for N = 240. Also the proportion of successful estimates is slightly lower but these are still acceptable proportions (see Table 7.3). Table 7.9 shows that there is some evidence that method PD is not performing as well with half the knots because the proportion of successful estimates is lower than that obtained for methods S and SAS which is not the case for n = 80 and N = 240 (see Table 7.3).

7.2 COMPARISON OF NON-PARAMETRIC AND PARAMETRIC METHODS

7.2.1 General comments

The methods discussed in Section 7.1 are non-parametric methods especially developed to handle non-linear calibration problems. The review in Chapter 2 showed that a high proportion of univariate calibration papers has been devoted to linear calibration. For example, Hunter and Lamboy's (1981a), Hoadley's (1970) and Aitchison and Dunsmore's (1975) approaches. This section will show that methods S and SAS can be usefully applied to data sets where the underlying

Table 7.8 Comparison of point estimates when assuming a uniform prior for ξ

Data set	Method	Mean absolute error	Calibration range
Gp40	S (Median)	0.42	0.0 – 13.26
	SAS (Median)	0.43	
	PD (Median)	0.38	
	PD (Mode)	0.46	
Wb40	S (Median)	0.87	10.00 – 79.52
	SAS (Median)	0.89	
	PD (Median)	0.85	
	PD (Mode)	0.88	

Fig.7.3: Comparison of point estimates
(Gp40 data set)

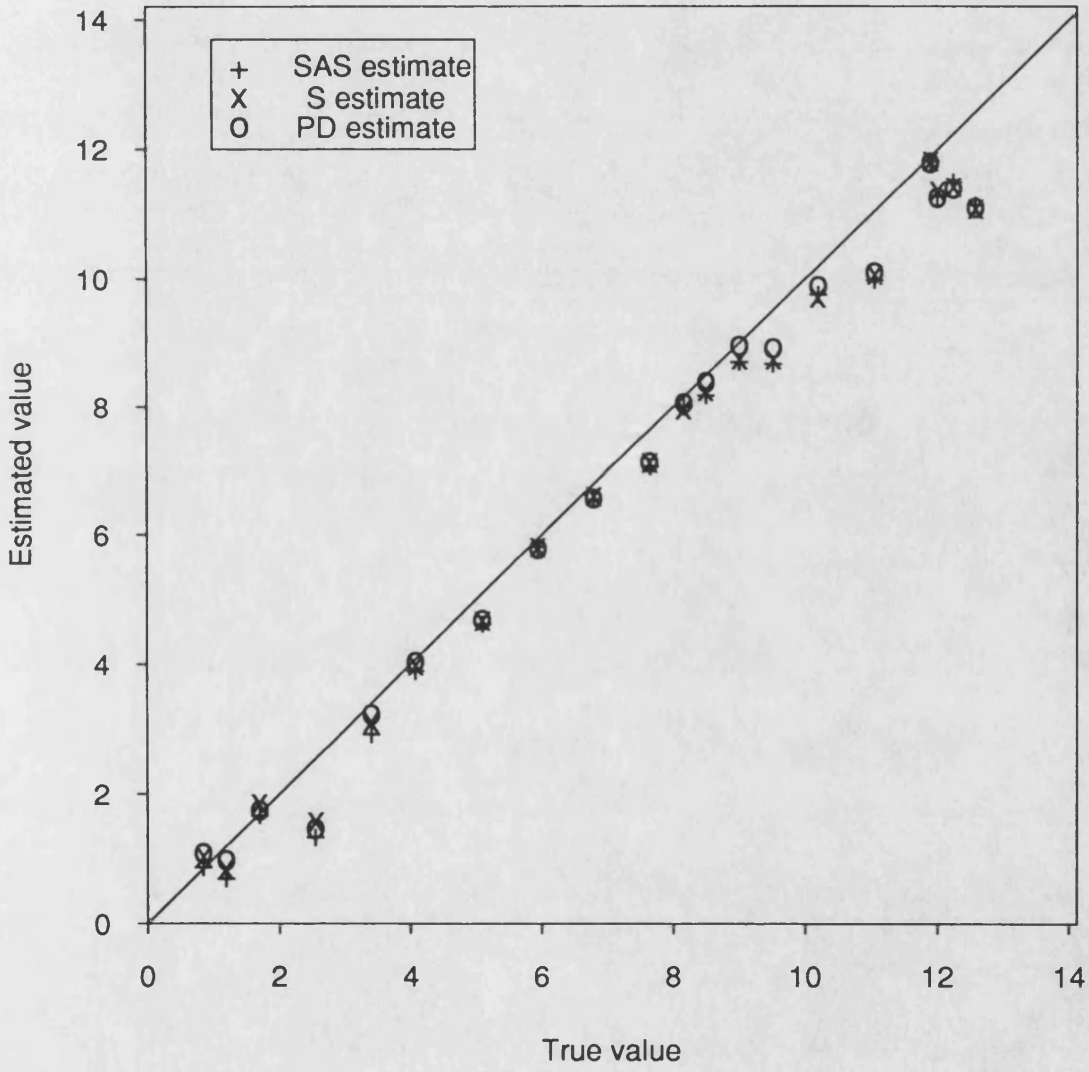


Table 7.9 Comparison of interval estimates when assuming a uniform prior for ξ

Data set	Method	Mean interval width	Proportion of successful estimates
Gp40	S	1.69	18/20
	SAS	1.67	18/20
	PD (SY)	1.69	18/20
	PD (HP)	1.67	18/20
Wb40	S	3.89	20/20
	SAS	3.86	20/20
	PD (SY)	4.04	19/20
	PD (HP)	3.93	19/20

model is believed to be linear. In the following comparisons, Hunter and Lamboy's, Hoadley's and Aitchison and Dunsmore's methods are denoted as HL, H and AD respectively.

In Chapters 3 and 4, methods S and SAS were applied not only to data sets arising from non-linear models (the Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder, Asymptotic and Antibiotic assay data sets) but also to a linear data set (Linear) and five data sets arising from measurements on teeth. Method S is a non-parametric method which possesses many similarities with method HL whereas method SAS is akin to methods H and AD because the latter all utilise prior information on ξ in the form of a prior distribution for ξ. This section compares the results obtained from applying the non-parametric methods S and SAS developed in this thesis with the established parametric methods HL, H and AD as follows:

(i) S and HL applied to the Linear and Weibull data sets.

(ii) S, SAS, H and AD applied to the Linear and teeth data sets.

Firstly let us briefly review methods HL, H and AD.

7.2.2 Methods HL, H and AD

Methods HL, H and AD are reviewed in detail in Chapter 2 (Section 2.3). With all these methods, the models for the calibration experiment and prediction stage are given by equations (2.1.1) and (2.1.2) respectively, i.e.

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i=1,2,\ldots,n \tag{7.2.1}$$

$$Y_j' = \beta_0 + \beta_1 \xi + \varepsilon_j' \qquad j=1,2,\ldots,m. \tag{7.2.2}$$

All the methods assume that the ε_i and ε_j' are i.i.d. normal random variables with mean zero and variance σ^2. Method HL assumed an *a priori* independence of (β_0,β_1) and η. For the case of σ^2 known, Hunter and Lamboy (1981a) assumed priors $p(\beta_0,\beta_1)$ and $p(\eta)$ were locally uniform whilst for the case of σ^2 unknown, they assumed a non-informative prior for $(\beta_0,\beta_1,\sigma^2,\eta)$ of the form

$$p(\beta_0,\beta_1,\sigma^2,\eta) \propto \sigma^{-2}.$$

With methods H and AD, prior information on ξ is directly utilised in the form of a prior density for ξ, $\pi(\xi)$, in contrast with method HL where the choice of prior for η *implies* a particular prior for ξ and prior information on ξ cannot easily be taken into account. Hoadley (1970) took a general form of prior density

$$p(\beta_0,\beta_1,\sigma^2,\xi) \propto p(\beta_0,\beta_1,\sigma^2)\pi(\xi)$$

with a non-informative prior for $(\beta_0,\beta_1,\sigma^2)$ given by $p(\beta_0,\beta_1,\sigma^2) \propto \sigma^{-2}$. Aitchison and Dunsmore (1975) assumed a non-informative prior for $(\beta_0+\beta_1\xi,\sigma^2)$. Both methods were based on a predictive density approach and the resulting posterior density of ξ is given by expression (7.1.3). The function $L(\xi)$ is given by expression (2.3.2) for method AD and by expression (2.3.3) for method H.

7.2.3 Comparison of methods S and HL applied to the Weibull and Linear data sets.

The Weibull data set was analysed using both non-parametric and parametric methods because of the five simulated data sets arising from non-linear models (Gompertz, Weibull, Preece-Baines, Bleasdale-Nelder and Asymptotic), the Weibull was the only data set where it was at all feasible to globally fit a straight line (see Fig. 3.8).

Methods S and HL were applied to the Linear and Weibull data sets and both point and interval estimates for ξ were obtained. Table 7.10 gives details of these estimates. Study of the table shows that for the Linear data set, method S gives better results than method HL. Part of the difference in the results from both methods is due to the fact that for method S, the posterior distribution of ξ is defined on $[x_1, x_n]$, i.e. it is zero outside the calibration range defined by $[x_1, x_n]$ whereas with method HL, the posterior distribution of ξ is, in theory, defined over the whole real line. Fig. 7.4 shows the estimated value of ξ, $\hat{\xi}$, plotted against the true value, ξ, for the Linear data set. The 45° line represents zero error and one can see from this plot that the point estimates obtained for both methods are fairly close to each other overall. As expected, Table 7.10 shows that for the Weibull data set, method HL is totally inadequate producing a very much larger mean absolute error, an unacceptably low proportion of successful estimates and less precise interval estimates for ξ than method S. Fig. 7.5 shows the estimated value, $\hat{\xi}$, plotted against the true value, ξ, for the Weibull data set, the 45° line representing zero error. It shows that method S gives very good results, whilst for method HL the errors are much greater and the pattern of errors reveals the inadequacy of a parametric method such as this for a non-linear data set.

7.2.4 Comparison of methods S, SAS, H and AD applied to the Linear and teeth data sets

The non-parametric methods S and SAS and the parametric methods H and AD were applied to the Linear data set. It was assumed for methods SAS, H and AD that ξ

Table 7.10 Comparison of methods HL and S assuming a uniform prior for ξ (Linear and Weibull data sets)

	Data set			
	Linear		Weibull	
Method	HL	S	HL	S
Mean absolute error	3.83	3.57	2.49	0.72
Mean interval width	15.80	15.13	6.60	3.53
Proportion of successful estimates	17/20	18/20	13/20	19/20
Calibration range	15.00 – 69.51		10.00 – 79.52	

Fig.7.4: Comparison of HL and S estimates
(Linear data set)

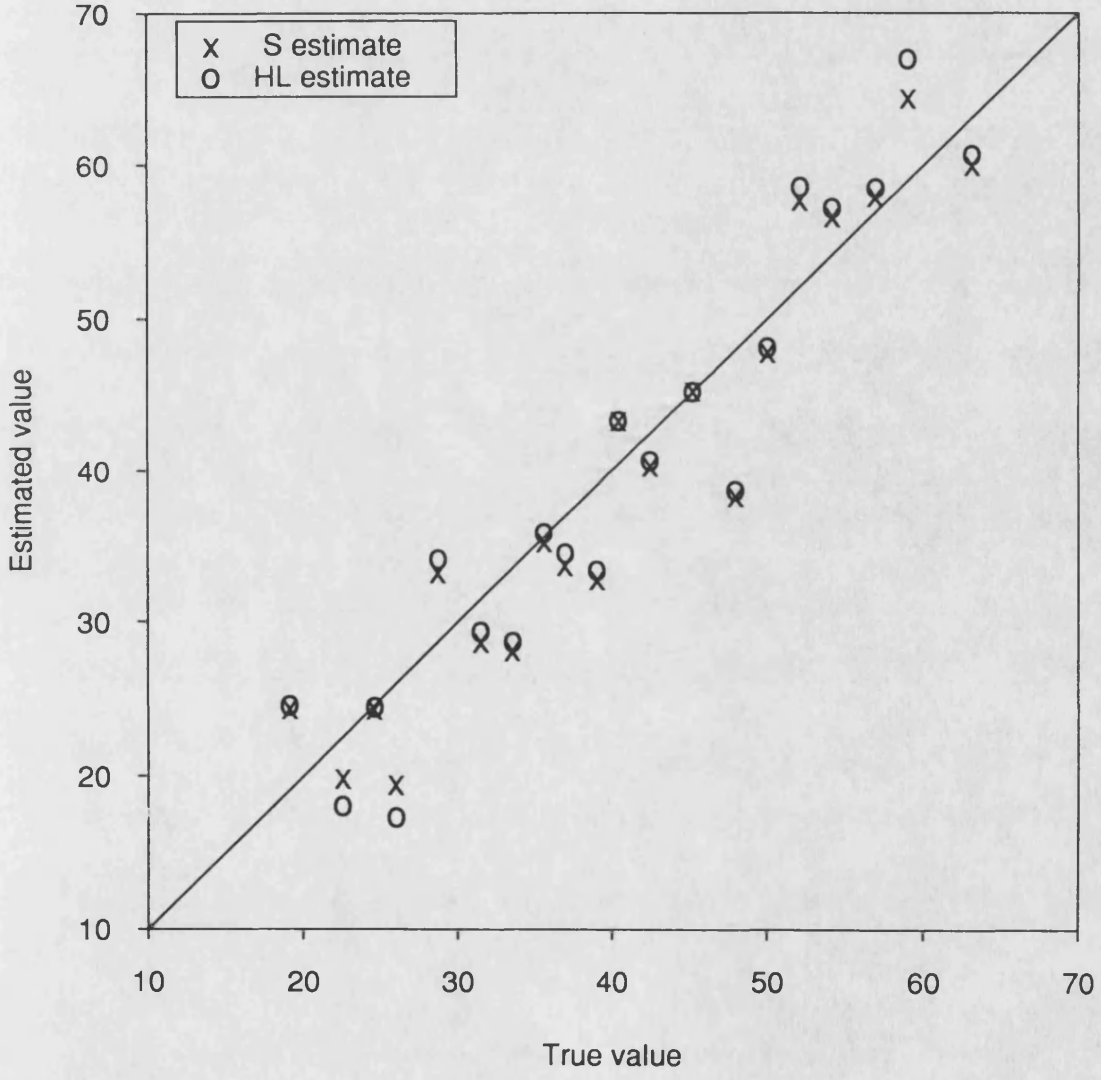
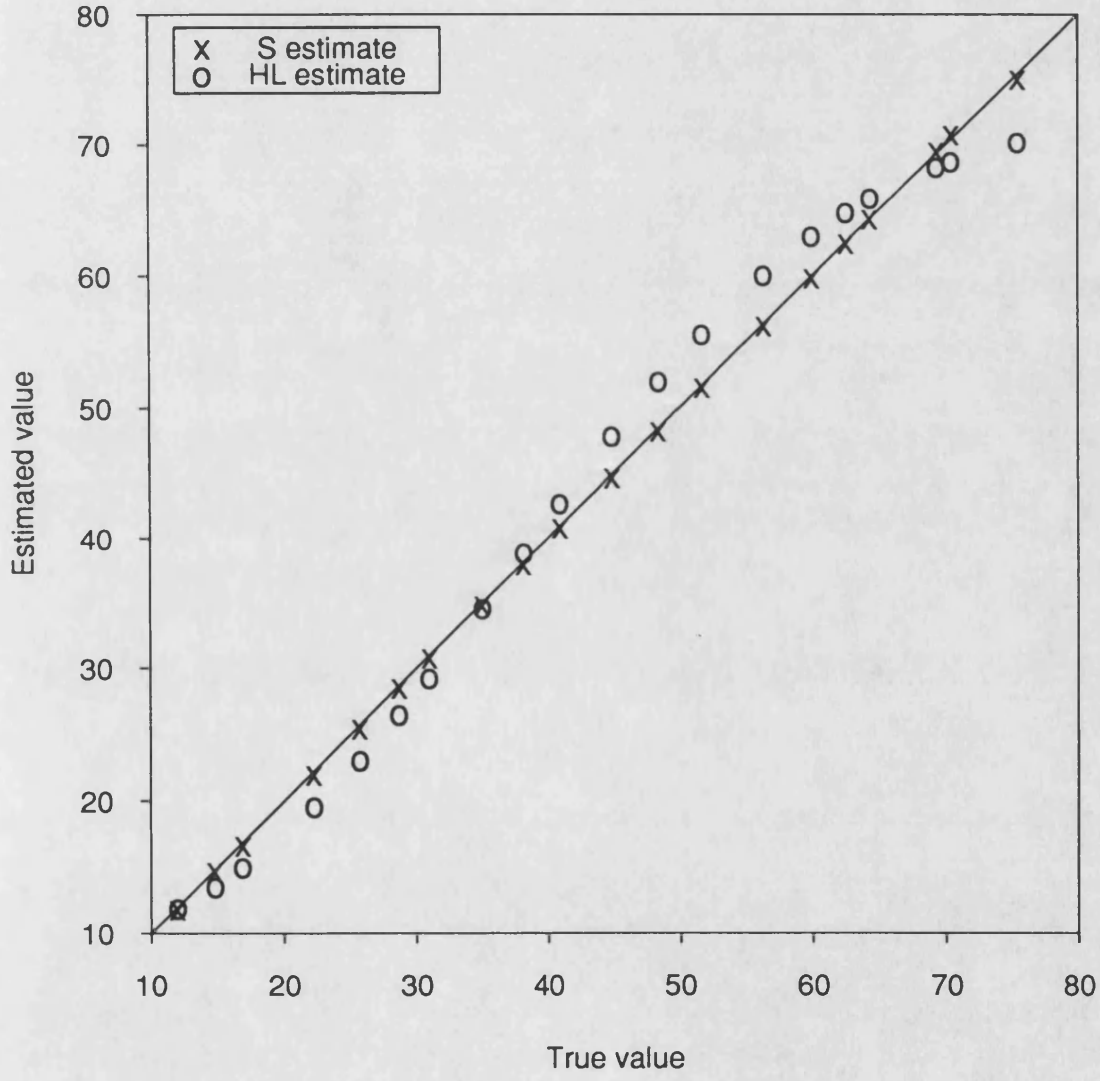


Fig.7.5: Comparison of HL and S estimates
(Weibull data set)



had a uniform prior density on $[x_1, x_n]$. Methods H and AD produced point and interval estimates identical to more than three decimal places. Fig. 7.6 shows the estimated value, $\hat{\xi}$, plotted against the true value, ξ, for the Linear data set using all four methods. The plot shows the point estimates to be extremely close. Tables 7.11 and Table 7.12 give details of the point and interval estimates for ξ obtained by applying all four methods to the Linear data set. It will be recalled that for methods S and SAS, the posterior median is used as a point estimator of ξ and the interval estimates for ξ are symmetrical. Table 7.11 gives the mean absolute errors for methods H and AD when using both the median and the mode as point estimators of ξ whereas Table 7.12 gives both highest posterior density (HP) and symmetrical (SY) interval estimates using methods H and AD. Non-parametric methods S and SAS compare very favourably with the parametric methods H and AD.

All four methods were also applied to the five teeth data sets (see Section 3.3.2, Chapter 3) and the results are compared in this section. A study of the scatter diagrams for all five teeth data sets showed that it was feasible to assume the true calibration curve was a straight line.

It will be recalled that with the teeth data, some patients contributed two teeth of a particular type, a right *and* a left tooth, whilst others contributed only one tooth of a particular type, a right *or* a left tooth. Obviously measurements of log ITTM on the right and left teeth belonging to the same patient are correlated. Methods S and SAS took this into account (see Section 3.3.2, Chapter 3). Suppose there are m patients with ages $x_1, x_2, \ldots, x_m$ each contributing one or two teeth of a particular type. For any particular type of tooth, the mathematical model for the calibration experiment is the following mixed model:

$$Y_{ik} = \beta_0 + \beta_1 x_i + \tau_i + \eta_{ik} \qquad \begin{matrix} i=1,2,\ldots,m \\ k=1 \text{ or } 2 \end{matrix}$$

where Y_{ik} = log ITTM for the kth tooth of the ith patient, τ_i is the effect due to the ith patient and β_0, β_1 are the intercept and slope parameters of the true calibration line. It is assumed that the τ_i are independent normally distributed random variables with mean zero and variance σ_1^2 and the η_{ik} are independent normally distributed random variables with mean zero and variance σ_2^2. It is further assumed that the η_{ik} are independent of the τ_i. Because of the lack of balance, the log likelihood function, $L(\theta)$, where $\theta = (\sigma_1^2\, \sigma_2^2\, \beta_0\, \beta_1)$, is very complicated. It is given in Osborne (1978). The maximum likelihood estimates of β_0, β_1 and σ^2 ($\sigma^2 = \sigma_1^2 + \sigma_2^2$) were obtained by maximising $L(\theta)$ using the Nelder-Mead simplex method (1965). It was not necessary to do this for the upper

Fig.7.6: Comparison of H, AD, S and SAS estimates
(Linear data set)

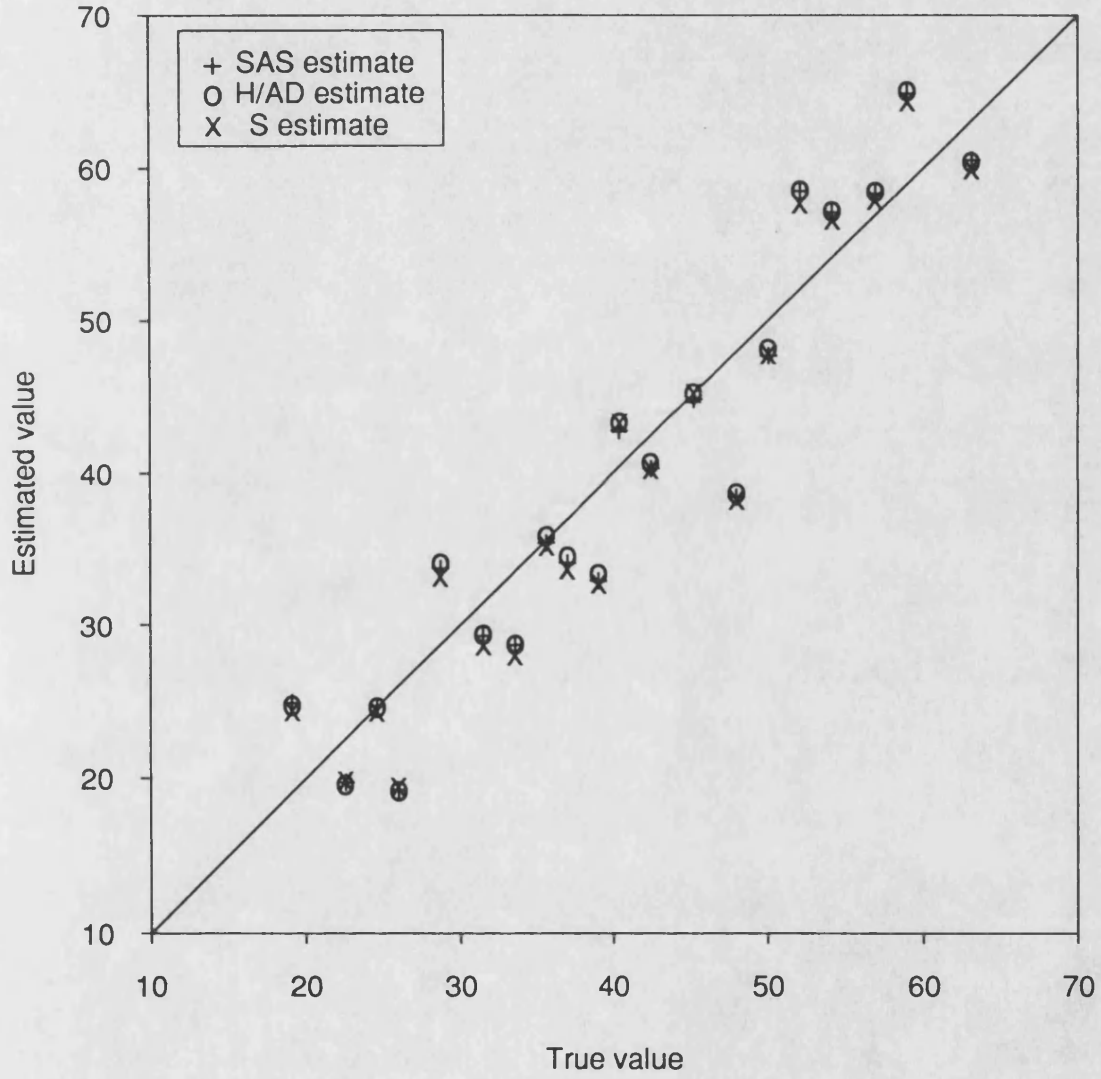


Table 7.11 Comparison of point estimates for the Linear data set

Method	Mean absolute error	Calibration range
S (Median)	3.6	15.00 – 69.51
SAS (Median)	3.6	
H/AD (Median)	3.6	
H/AD (Mode)	3.8	

Table 7.12 Comparison of interval estimates for the Linear data set (90% probability level)

Method	Mean interval width	Proportion of successful estimates
S	15.1	18/20
SAS	15.3	19/20
H/AD (SY)	14.8	19/20
H/AD (HP)	14.6	17/20

canine calibration data set because it was decided to assume independence of observations as this data set had certain peculiarities (Osborne, 1978). Methods H and AD used the respective estimates of β_0, β_1 and σ^2, to produce the posterior distributions of ξ and for methods SAS, H and AD a uniform prior density on $[x_1, x_n]$ was assumed. Methods H and AD produced both point and interval estimates for ξ which were identical to more than three decimal places. Fig. 7.7 shows the estimated age of a *new* patient, $\hat{\xi}$, plotted against the true age of the patient, ξ, for all four methods applied to the upper central teeth data, the 45° line representing zero error. For most of the new patients, the estimates of age obtained using the four methods are reasonably close. Table 7.13 gives details of the point estimates of ξ which were obtained. A comparison of the mean absolute errors, using the posterior median as a point estimator of ξ, shows that the methods are very comparable. Table 7.14 gives details of the interval estimates for ξ obtained using all four methods. Again both highest posterior density (HP) and symmetrical interval estimates (SY) were obtained with methods H and AD. A study of this table shows that overall the non-parametric methods give slightly better results.

7.3 CONCLUSIONS

The non-parametric methods S (simulation), SAS (switching algorithm and simulation) and PD (predictive density) developed in Chapters 3, 4 and 5 respectively all produced good point and interval estimates for ξ when applied to a wide variety of simulated and real data sets. With reference to point estimation, mean absolute errors were small compared to the calibration range and all the methods possessed a good property that where the data were less informative concerning the value of ξ (i.e. the spline was flatter) errors were larger. All three methods produced interval estimates for ξ which were acceptable with respect to both precision and proportion of successful estimates. There seemed to be no obvious disadvantages with any of the methods and all were easy to implement.

Of the three developed methods, S, SAS and PD, the most versatile would seem to be method SAS as this produces good results for all the data sets used in Chapters 3, 4 and 5 and it is equally useful in both the situation where f is a cubic smoothing spline and the situation where f is a cubic interpolating spline. It would be good to consider extending this method to apply to multivariate calibration problems by using, for example, additive modelling.

Methods S and SAS are very flexible because they can be applied to any data set even if the underlying model is believed to be linear and also to data sets

Fig.7.7: Comparison of H, AD, S and SAS estimates

(Upper central teeth data set)

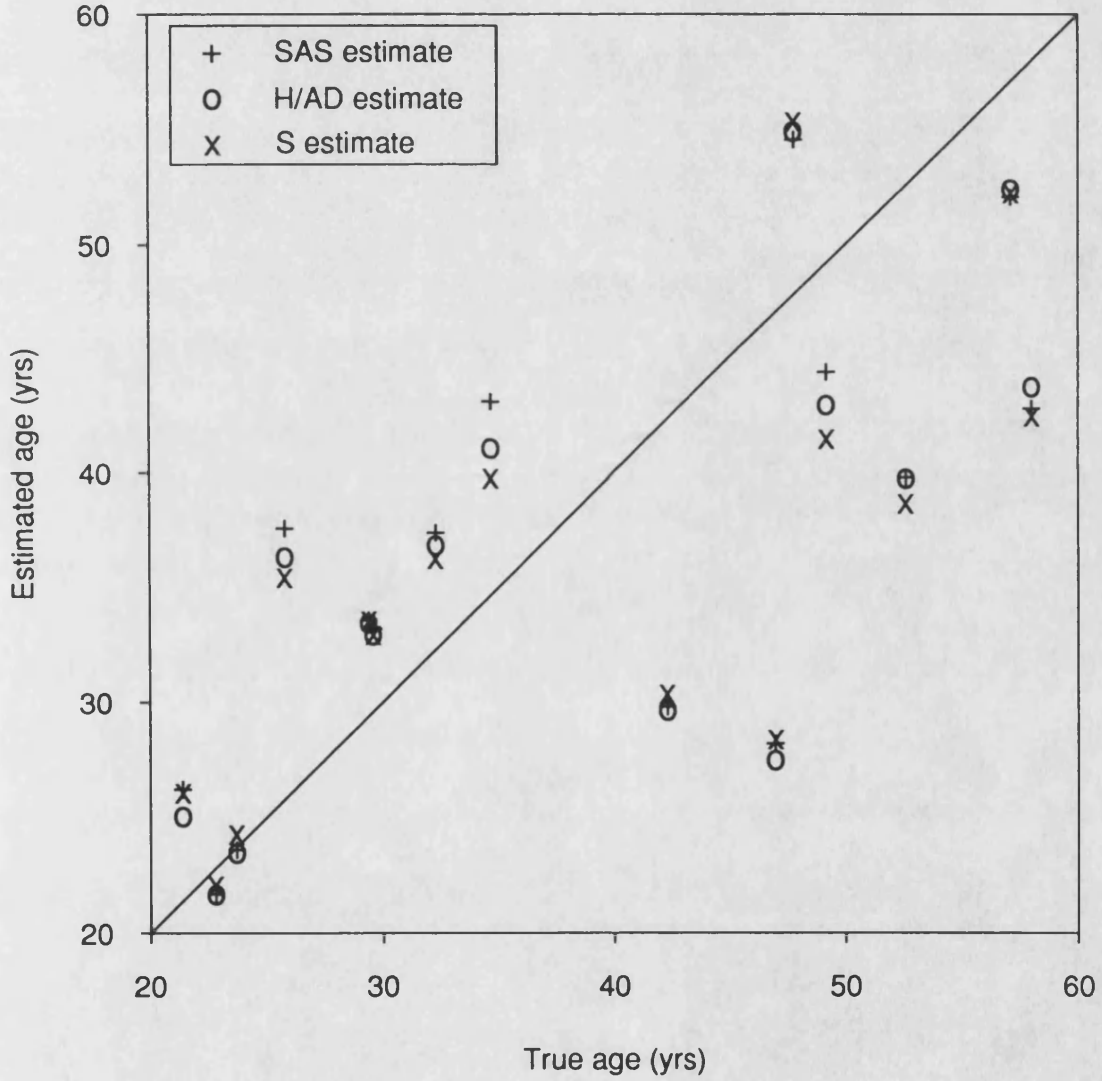


Table 7.13 Comparison of point estimates for the teeth data sets

Data set	Method	Mean absolute error	Calibration range
Upper central	S (Median)	7.5	16.1 – 62.5
	SAS (Median)	7.7	
	H/AD (Median)	7.4	
	H/AD (Mode)	8.2	
Upper canine	S (Median)	4.6	20.3 – 62.5
	SAS (Median)	4.7	
	H/AD (Median)	4.6	
	H/AD (Mode)	6.5	
Lower central	S (Median)	5.7	26.6 – 62.5
	SAS (Median)	5.6	
	H/AD (Median)	5.8	
	H/AD (Mode)	7.0	
Lower lateral	S (Median)	5.6	27.8 – 62.5
	SAS (Median)	5.7	
	H/AD (Median)	8.7	
	H/AD (Mode)	8.8	
Lower canine	S (Median)	10.1	22.8 – 65.6
	SAS (Median)	9.8	
	H/AD (Median)	9.7	
	H/AD (Mode)	11.6	

Table 7.14 Comparison of interval estimates for the teeth data sets (80% probability level)

Data set	Method	Mean interval width	Proportion of successful estimates
Upper central	S	17.1	10/15
	SAS	17.7	10/15
	H/AD (SY)	18.2	10/15
	H/AD (HP)	17.7	9/15
Upper canine	S	15.3	5/7
	SAS	15.0	5/7
	H/AD (SY)	17.0	6/7
	H/AD (HP)	15.0	6/7
Lower central	S	15.2	3/3
	SAS	14.0	2/3
	H/AD (SY)	17.8	3/3
	H/AD (HP)	15.6	2/3
Lower lateral	S	20.6	5/6
	SAS	20.2	5/6
	H/AD (SY)	24.5	5/6
	H/AD (HP)	21.6	5/6
Lower canine	S	20.0	9/13
	SAS	19.6	9/13
	H/AD (SY)	21.8	9/13
	H/AD (HP)	20.2	9/13

where the observations are proportions, both Binomial and non-Binomial. In Chapter 6, by way of illustration, method S was applied to both real and simulated data sets where the observations were proportions and gave very pleasing results.

The comparison of results in Section 7.1. would seem to suggest certain recommendations. The first two recommendations apply to the situation where it can be assumed that the true calibration curve, f, is a monotonic natural cubic *smoothing* spline with knots at the data points $\{x_i\}$ $i=1,2,\ldots,n$.

(i) If there is no prior information about ξ or only vague prior information on ξ, use method S which views ξ as a non-linear functional of f and uses simulation to obtain the posterior distribution of ξ.

(ii) If there is definite prior information on ξ which can be expressed in the form of a prior density for ξ which is unimodal on $[x_1, x_n]$, then use method SAS which again views ξ as a non-linear functional of f and uses a switching algorithm together with simulation to produce the posterior distribution of ξ.

When using method S or SAS, symmetrical interval estimates can be readily obtained from the posterior distribution of ξ and the posterior median is a good point estimator of ξ.

The last recommendation, (iii), applies to where it can be assumed that f is a natural cubic *interpolating* spline with knots at the data points $\{x_i\}$ $i=1,2,\ldots,n$.

(iii) Use method PD, which is based on predictive densities, to obtain the posterior distribution of ξ. The posterior mode is a good point estimator of ξ and highest posterior density interval estimates for ξ can be easily obtained from the posterior distribution of ξ.

With reference to methods S, SAS and PD developed in this thesis, the only comparable approach indicated by the review in Chapter 2, is given in a recent paper by Racine-Poon (1988). It would be of interest in future to apply methods S, SAS and PD to the assay data sets given in this paper and compare the results obtained with those calculated by Racine-Poon.

BIBLIOGRAPHY

Abramowitz, M. & Stegun, I.A. (1965). *Handbook of mathematical functions.* National Bureau of Standards. Applied mathematics series. **55.**

Aitchison, J. & Dunsmore, I.R. (1975). *Statistical Prediction Analysis.* Cambridge University Press.

Ali, M.A. & Singh, N. (1981). An alternative estimator in inverse linear regression. *J. Statist. Comp.* **14**, 1-15.

Atkinson, A.C. & Whittaker, J. (1976). A switching algorithm for the generation of beta random variables with at least one parameter less than one. *J. R. Statist. Soc. A,* **139**, 462-467.

Barnett, V.D. (1966). Discussion of a paper by Sprent. *J. R. Statist. Soc. B,* **28**, 291-292.

Barnett, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics* **25**, 129-142.

Berkson, J. (1969). Estimation of a linear function for a calibration line : consideration of a recent proposal. *Technometrics* **11**, 649-660.

Box, G.E.P. & Tiao, G.C. (1973). *Bayesian inference in statistical analysis.* Addison-Wesley.

Brown, G.H. (1979). An optimisation criterion for linear inverse estimation. *Technometrics,* **21**, 575-579.

Brown, P.J. (1982). Multivariate calibration (with discussion). *J. R. Statist. Soc. B,* **44**, 287-321.

Brown, P.J. & Sundberg, R. (1987) Confidence and conflict in multivariate calibration. *J. R. Statist. Soc. B,* **49**, 46-57.

Brown, P.J. & Sundberg, R. (1989). Multivariate calibration with more variables than observations. *Technometrics,* **31**, 365-371.

Carroll, R.J., & Spiegelman, C.H. (1988). A quick and easy multiple-use calibration-curve procedure. *Technometrics,* **30**, 137-141.

Chow, S.-C. & Shao, J. (1990). On the difference between the classical and inverse methods of calibration. *Applied Statistics,* **39**, 219-228.

Clark, R.M. (1979). Calibration, Cross-validation and carbon - 14. I. *J. R. Statist. Soc. A,* **142**, 47-62.

Clark, R.M. (1980). Calibration, Cross-validation and carbon - 14.II. *J. R. Statist. Soc. A,* **143**, 177-194.

Cochran, W.C. (1943). The comparison of different scales of measurement for experimental results. *Ann. Math. Statist.*, **14**, 205-216.

Copas, J.B. (1982). Discussion of P.J. Brown's paper. *J. R. Statist. Soc. B*, **44**, 312-313.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31**, 377-403.

Creasy, M.A. (1954). Limits for the ratio of means. *J. R. Statist. Soc. B*, **16**, 186-194.

De Boor, C. (1978). *A practical guide to splines*. Springer-Verlag.

De Doncker, E. (1978). An adaptive extrapolation algorithm for automatic integration. *Signum newsletter*, **13, 2**, 12-18.

Dunsmore, I.R. (1968). A Bayesian approach to calibration. *J. R. Statist. Soc. B*, **30**, 396-405.

Eagleson, G.K. & Buckley, M.J. (1988). An approximation to the distribution of quadratic forms in normal random variables. *Austral. J. Statist.*, **30A**, 150-159.

Easterling, R.G. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics*, **23**, 343-344.

Edgeworth, F.Y. (1898). *J. R. Statist. Soc.*, **61**, 670

Edwards, A.W.F. (1972). *Likelihood.* Cambridge University Press.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Statist.*, **10**, 162-186.

Eubank, R.L. (1988). *Spline smoothing and non-parametric regression.* Marcel Dekker.

Fairfield-Smith, H. (1936). The problem of comparing the results of two experiments with unequal errors. *J. Council Scientific & Industrial Research*, **9**, 211-212.

Farebrother, R.W. (1984). A note on Fearn's "Misuse of ridge regression". *Applied Statistics*, **33**, 74-75.

Fearn, T. (1983). Misuse of ridge regression in the calibration of a near infrared reflectance instrument. *Applied Statistics*, **32**, 73-79.

Fieller, E.C. (1954). Some problems in interval estimation. *J. R. Statist. Soc. B*, **16**, 175-185.

Fujikoshi, Y. & Nishii, R. (1984). On the distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Math. J.*, **14**, 215-225.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika,* **38**, 481-482.

Gill, P.E. & Miller, G.F. (1972). An algorithm for the integration of unequally spaced data. *Comp. Journal.* **15**, 80-83.

Gill, P.E. & Murray, W. (1976). Minimisation subject to bounds on the variables. *National Physical Laboratory report NAC 72.*

Greville, T.N.E. (ed.) (1969). *Theory and application of spline functions.* Academic Press.

Grubbs, F.E. (1948). On estimating precision of measuring instruments and product variability. *J. Amer. Statist. Ass.,* **43**, 243-264.

Grubbs, F.E. (1973). Errors of measurement, precision, accuracy, and the statistical comparison of measuring instruments. *Technometrics,* **15**, 53-66.

Halperin, M. (1970). On inverse estimation in linear regression. *Technometrics,* **12**, 727-736.

Hill, B.M. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics,* **23**, 335-338.

Hoadley, B. (1970). A Bayesian look at inverse linear regression. *J. Amer. Statist. Ass.,* **65**, 356-369.

Hoerl, A.E. & Kennard, R.W. (1970). Ridge regression : biased estimation for non orthogonal problems. *Technometrics,* **12**, 55-67

Hoerl, A.E., Kennard, R.W. & Hoerl, R.W. (1985). Practical use of ridge regression : a challenge met. *Applied Statistics,* **34**, 114-120.

Hunter, W.G. & Lamboy, W.F. (1981a). A Bayesian analysis of the linear calibration problem. *Technometrics,* **23**, 323-328.

Hunter, W.G. and Lamboy, W.F. (1981b). Response to discussion of their 1981 paper. *Technometrics,* **23**, 344-350.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika,* **48**, 419-426.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika,* **36**, 149-176.

Johnson, N.L. & Kotz, S. (1970). *Continuous univariate distributions. Vol. 2.* Houghton Mifflin.

Kalbfleisch, J.D. & Sprott, D.A. (1970). Applications of likelihood methods to models involving large numbers of parameters (with discussion). *J. R. Statist. Soc., B,* **32**, 175-208.

Kalotay, A.J. (1971). Structural solution to the linear calibration problem. *Technometics,* **13**, 761-769.

Kimeldorf, G. and Wahba, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, **41**, *495-502*.

Kimeldorf, G. and Wahba, G. (1970b). Spline functions and stochastic processes. *Sankhya (A)*, **32**, 173-180.

Knafl, G., Spiegelman, C., Sacks, J. & Ylvisaker, D. (1984). Nonparametric calibration. *Technometrics*, **26**, 233-241.

Krutchkoff, R.G. (1967). Classical and inverse regression methods of calibration. *Technometrics*, **9**, 425-439.

Krutchkoff, R.G. (1968). Letter to the editor. *Technometrics*, **10**, 430-431.

Krutchkoff, R.G. (1969). Classical and inverse regression methods of calibration in extrapolation. *Technometrics*, **11**, 605-608.

Krutchkoff, R.G. (1971). The calibration problem and closeness. *J. Statist. Comput. Simul. I*, 87-95.

Lawless, J.F. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics*, **23**, 334-335.

Lechner, J., Reeve, C. & Spiegelman, C. (1982). An implementation of the Scheffé approach to calibration using spline functions, illustrated by a pressure-volume calibration. *Technometrics*, **24**, 229-234.

Lehmann , E.L. (1975) *Nonparametrics : Statistical methods based on ranks*. McGraw-Hill.

Li, K.C. (1982). Minimaxity of the method of regularisation on stochastic processes. *Ann. Statist.* **10**, 937-942.

Lieberman, G.J., Miller, R.G. & Hamilton, M.A. (1967). Unlimited simultaneous discrimination intervals in regression. *Biometrika*, **54**, 133-145.

Lieftinck-Koeijers, C.A.J. (1988) Multivariate calibration: a generalisation of the classical estimator. *J. Multivariate Anal.* **25**, 31-44.

Lindley, D.V. (1972). *Bayesian Statistics : A Review*. Philadelphia : SIAM.

Linnig, F.J. & Mandel, J. (1964). Which measure of precision? *Analy. Chem.*, **36**, 25A-32A

Lundberg, E. & De Mare, J. (1980). Interval estimates in the spectroscopy calibration problem. *Scand. J. Statist.*, **7**, 40-42.

Lwin, T. (1981). Discussion of Hunter & Lamboy's paper (1981). *Technometrics*, **23**, 339-341.

Lwin, T. & Maritz, J.S. (1980). A note on the problem of statistical calibration. *Applied Statistics*, **29**, 135-141.

Lwin, T. & Maritz, J.S. (1982). An analysis of the linear calibration controversy from the perspective of compound estimation. *Technometrics,* **24**, 235-242.

Lwin, T. and Spiegelman, C.H. (1986). Calibration with working standards. *Applied Statistics,* **35**, No. 3, 256-261.

Mandel, J. (1958). A note on confidence intervals in regression problems. *Ann. Math. Stats.,* **29**, 903-907.

Mandel, J. (1984). Fitting straight lines when both variables are subject to error. *J. Qual. Tech.* **16**, 1-4.

Mandel, J. & Linnig, F.J. (1957). Study of accuracy in chemical analysis using linear calibration curves. *Analy. Chem.,* **29**, 743-749.

Martinelle, S. (1970). On the choice of regression in linear calibration. *Technometrics,* **12**, 157-161.

McCullagh, P. & Nelder, J.A. (1989). *Generalised linear models.* Chapman and Hall.

Miller, R.G. (1966). *Simultaneous Statistical Inference.* McGraw-Hill.

Minder, Ch. E. & Whitney, J.B. (1975). A likelihood analysis of the linear calibration problem. *Technometrics,* **17,** 463-471.

Miwa, T. (1985). Comparison among point estimators in linear calibration in terms of mean squared error. *Jp. J. Appl. Statist.,* **14**, 83-93.

Morgan, B.J.T. (1984). *Elements of Simulation.* Chapman and Hall.

Muhammad, F. (1987). Linear statistical calibration. *Ph. D. thesis, University of Glasgow.*

Naes, T. (1985a). Comparison of approaches to multivariate linear calibration. *Biom. J.,* **27**, 265-275.

Naes, T. (1985b). Multivariate calibration when error covariance-matrix is structured. *Technometrics,* **27**, 301-311.

Naes, T. (1986). Multivariate calibration using covariance adjustment. *Biom. J.,* **28**, 99-107.

Naes, T., Irgens, C., & Martens, H. (1986). Comparison of linear statistical methods for calibration of NIR instruments. *Applied Statistics,* **35,** 195-206.

Naes, T. & Martens, H. (1984). Multivariate calibration. II. Chemometric Methods. *Trends. Anal. Chem.,* **3**, 266-271.

Naszodi, L.J. (1978). Elimination of the bias in the course of calibration. *Technometrics,* **20**, 201-206.

Naylor, J.C. & Smith, A.F.M. (1982) Applications of a method for the efficient computation of posterior distributions. *Applied Statistics,* **31**, 214-225.

Nelder, J.A. & Mead, R. (1965). A simplex method for function minimisation. *Comp. J.* **7**, 308-313.

Nishii, R. & Krishnaiah, P.R. (1988). On the moments of classical estimates of explanatory variables under a multivariate calibration model. *Sankhya,* **50**, 137-148.

Oden, A. (1973). Simultaneous confidence intervals in inverse linear regression. *Biometrika,* **60**, 339-343.

Oman, S.D. (1985). An exact formula for the M.S.E. of the inverse estimator in linear calibration problem. *J. Statist. Planning Infer.,* **11**, 189-196.

Oman, S.D. (1988). Confidence regions in multivariate calibration. *Ann. Statist.,* **16**, 174-187.

Oman, S. & Wax, Y. (1984). Estimating foetal age by ultrasound measurements : an example of multivariate calibration. *Biometrics,* **40**, 947-960.

Orban, J.E. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics,* **23**, 342-343.

Osborne, C. (1978). Two statistical analyses : (1) Transmitter release at neuromuscular junctions, (2) Dental estimation of age. *M. Sc. thesis, University of Glasgow.*

Perng, S.K. (1987). A note on the inverse estimator for the linear calibration problem. *Commun. Statist. A.* **16**, 1743-1747.

Perng, S.K. & Tong, Y.L. (1974). Sequential solution to inverse linear regression problem. *Ann. Statist.* **2**, 535-9.

Perng, S.K. & Tong, Y.L. (1977). Optimal allocation of observations in inverse regression — note. *Ann. Statist.* **5**, 191-6.

Piessens, R., De Doncker-Kapenga, E., Uberhuber, C. & Kahaner, D. (1983). *A subroutine package for automatic integration.* Springer-Verlag.

Prudnikov, A.P., Brychkov, Yu. A. & Marichev, O.I. (1986). *Integrals and series: Vol.1. Elementary functions.* Gordon and Breach.

Racine-Poon, A. (1988). A Bayesian approach to non-linear calibration problems. *J. Amer. Statist. Ass.,* **83**, 650-656.

Ratkowsky, D.A. (1983). *Nonlinear regression modeling.* Marcel Dekker.

Reinsch, C. (1967). Smoothing by spline functions I. *Numer. Math.,* **10** 177-183.

Reinsch, C. (1971). Smoothing by spline functions II. *Numer. Math.,* **16**, 451-454.

Rosenblatt, J.R. & Spiegelman, C.H. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics,* **23**, 329-333.

Rothman, D. (1968). Letter to the editor. *Technometrics,* **10**, 429-430.

Satterthwaite, F.E. (1941). Synthesis of variance. *Psychometrika,* **6**, 309-316.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics,* **2**, 110-114.

Saw, J.G. (1970). Letter to the editor. *Technometrics,* **12**, p.937.

Sayers, M.W., Gillespie, T.D. & Queiroz, C.A.V. (1986a). The International Road Roughness Experiment : Establishing Correlation and a Calibration Method for Measurements. *Technical Paper 45, World Bank, Washington D.C.*

Scheffé, H. (1959). *The analysis of variance*. Wiley.

Scheffé, H. (1973). A statistical theory of calibration. *Ann. Statistics,* **1**, 1-37.

Schoenberg, I.J. (1964). Spline function and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.,* **52**, 947-950.

Shukla, G.K. (1972). On the problem of calibration. *Technometrics,* **14**, 547-553.

Shukla, G.K. & Datta, P. (1985). Comparison of the inverse estimator with the classical estimator subject to a preliminary test in linear calibration. *J. Statist. Planning Infer.,* **12**, 93-102.

Silverman, B.W. (1982). Kernel density estimation using the fast Fourier transform. *Applied Statistics,* **31**, 93-99.

Silverman, B.W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Amer. Statist. Ass.,* **79**, 584-589.

Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc. B.,* **47**, 1-52.

Silverman, B.W. & Watters, G. (1984) BATHSPLINE. *An interactive spline smoothing package. School of Mathematical Sciences*, University of Bath.

Smith, R.L. & Corbett, M. (1987). Measuring marathon courses: an application of statistical calibration theory. *Applied Statistics,* **36**, 283-295.

Solomon, H. & Stephens, M.A. (1977). Distribution of a weighted sum of chi-square variables. *J. Amer. Statist. Ass.,* **72**, 881-885.

Speckman, P. (1989). The asymptotic integrated mean square error for smoothing noisy data by splines. *Numer. Math.*, (to appear).

Spezzaferri, F. (1985). A note on multivariate calibration experiments. *Biometrics,* **14**, 267-272.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with Discussion). *J. R. Statist. Soc., B,* **36**, 111-147.

Stone, M. (1977). Consistent non-parametric regression. *Ann. Statist.*, **5**, 595-645.

Sundberg, R. (1985). When is the inverse regression estimator M.S.E. superior to the standard regression estimator in multivariate controlled calibration situations? *Stat. Prob. L.*, **3**, 75-79.

Tallis, G.M. (1969). Note on a calibration problem. *Biometrika,* **56**, 505-508.

Theobald, C.M. & Mallinson, J.R. (1978). Comparative calibration, linear structural relationships and congeneric measurements. *Biometrics,* **34**, 39-45.

Turiel, T.P., Hahn, G.J. & Tucker, W.T. (1982). New simulation results for the calibration and inverse median estimation problems. *Commun. Statist.-Simul. Comp.,* **11**, 677-713.

Vecchia, D.F., Iyer, H.K. & Chapman, P.L. (1989). Calibration with randomly changing standard curves. *Tehnometrics,* **31**, 83-90.

Wahba, G. (1975). Smoothing noisy data with spline functions. *Numer. Math.,* **24,** 383-393.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. J. R. Statist. Soc. B, **49**, 364-372.

Wegman, E.J. & Wright, I.W. (1983) Splines in statistics. *J. Amer. Statist. Ass.,* **78,** 351-365.

Welch, B.L. (1936). The specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. *J. R. Statist. Soc. Supp.* **3**, 29-48.

Welch, B.L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika,* **39**, 350-361.

Williams, E.J. (1959). *Regression Analysis.* John Wiley.

Williams, E.J. (1969a). Regression methods in calibration problems. *Bull. Intern. Statist. Inst.,* **43**, 17-28.

Williams, E.J. (1969b). A note on regression methods in calibration. *Technometrics,* **11,** 189-192.

APPENDIX 1. DETAILS OF DATA SETS.

A1.1 THE SIMULATED DATA SETS

The non-parametric methods described in Chapters 3, 4 and 5 were assessed by applying them to six simulated data sets, five simulated from non-linear models (Gompertz, Weibull, Bleasdale-Nelder, Asymptotic and Preece-Baines) and one from a linear model (Linear).

Ratkowsky (1983) calls the Gompertz and Weibull models *sigmoidal growth* models. Processes which produce sigmoidal or S-shaped growth curves are widespread in biology, agriculture, economics and engineering. Such curves start at some fixed point and increase their growth rate monotonically to reach an inflexion point; after this the growth rate decreases to approach asymptotically some final value. The equations of the models together with the values of the parameters used in the simulation are as follows

$$\textit{Gompertz} \quad Y = \alpha \exp[-\exp(\beta - \gamma X)] \tag{A1.1}$$

with $\alpha = 22.51$, $\beta = 2.106$ and $\gamma = 0.388$.

$$\textit{Weibull} \quad Y = \alpha - \beta \exp(-\gamma X^{\delta}) \tag{A1.2}$$

with $\alpha = 70.0$, $\beta = 61.7$, $\gamma = 0.0001$ and $\delta = 2.4$.

The Bleasdale-Nelder model is termed by Ratkowsky (1983) a *yield-density* model since it is appropriate to use where Y is the yield of a crop and X is the spacing or density of planting of the crop. Its equation is given by

$$Y = (\alpha + \beta X)^{-\frac{1}{\theta}} \tag{A1.3}$$

The values of α, β and θ used for the simulation were 0.02, 0.0003 and 13.94 respectively.

The Asymptotic model has been used extensively in agriculture and to a lesser extent in biology and the engineering sciences (Ratkowsky, 1983). The model is similar to the sigmoidal growth models in that the curve approaches an asymptote as X increases, but differs in that it lacks an inflection point and hence is not sigmoidal in shape. The equation defining the model is given by

$$Y = \alpha - \beta\gamma^{X} \tag{A1.4}$$

The values of α, β and γ chosen for the simulation were 2.7, 0.97 and 0.87 respectively.

The Preece-Baines model was one of several models suggested by Preece and Baines (1978). The model was put forward as being suitable for modelling

the growth of children where Y is the height of the child at a given time and X is the age at the child at that time. Its equation is given by

$$Y = \frac{4(h_1 - h_\theta)}{\{\exp[p_0(X-\theta)] + \exp[p_1(X-\theta)]\}\{1+\exp[q_1(X-\theta)]\}} \quad \text{(A1.5)}$$

Here h_1 is the final (or adult) height in cms. The values of the parameters used for the simulation are given by $h_1 = 174.0$, $h_\theta = 164.0$, $p_0 = 0.0880$, $p_1 = 0.2245$, $q_1 = 1.3676$ and $\theta = 14.75$.

Finally the sixth model was a linear model given by

$$Y = 0.08 + 0.015X \quad \text{(A1.6)}$$

Suppose Y_s denotes a simulated value from one of the six simulated data sets used in Chapters 3, 4 and 5, then $Y_s = Y + \varepsilon_s$ where Y is given by one of the equations $(A1.1) - (A1.6)$ and it is assumed that ε_s is a normally distributed random error term with mean 0 and variance σ^2. A total of eighty equally-spaced values of X were chosen to cover a suitable range. The values of σ and the range of X values used in the simulations were as follows

Gompertz:	$\sigma = 1.024$	$0.0 \le X \le 13.43$
Weibull:	$\sigma = 1.300$	$10.0 \le X \le 79.52$
Bleasdale–Nelder:	$\sigma = 4.252$	$20.0 \le X \le 178.0$
Asymptotic:	$\sigma = 0.080$	$1.00 \le X \le 21.74$
Preece–Baines:	$\sigma = 1.760$	$4.00 \le X \le 19.80$
Linear:	$\sigma = 0.128$	$15.00 \le X \le 69.51$

A1.2 THE ANTIBIOTIC ASSAY DATA SET

It is extremely important that doctors can estimate reasonably accurately the concentration of antibiotics, such as Tobramycin, in the blood of patients because if the level of antibiotic reaches more than 10 μg/ml even once in a day, it may deafen or kill a child patient. If a specified volume of an antibiotic is placed on an infected medium in a Petri dish and kept under controlled conditions, a circular area of medium is cleared by the antibiotic. The diameter of the cleared area, called the *clearance diameter*, depends on the level of concentration of antibiotic.

Let us consider the calibration procedure for such a biological assay. The calibration experiment consists of measuring many clearance diameters (mms) corresponding to *known* concentrations of an antibiotic (μg/ml) dropped into a batch of infected medium. The calibration data can be used to estimate the

calibration or response curve. The idea is then to take a specified volume of the patient's blood containing an *unknown* concentration of antibiotic (ξ μg/ml), drop it into the *same* batch of infected medium, observe the clearance diameter (D') and then estimate ξ given the estimated response curve and D'. Obviously several measurements D' could be made at the prediction stage of this calibration procedure if desired.

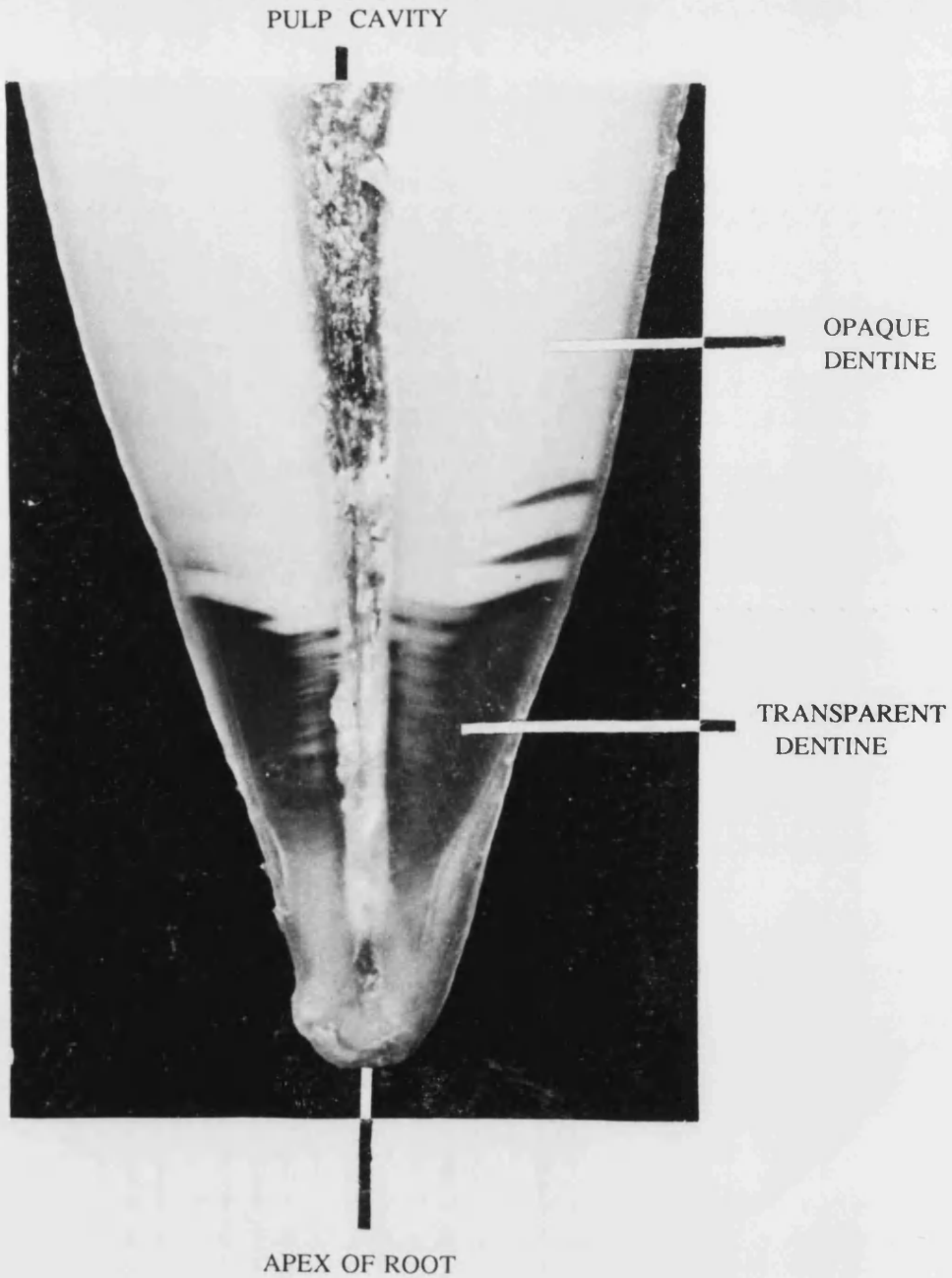
The data set analysed in Chapters 3, 4 and 5 of the thesis was obtained from a Glasgow Hospital and is similar to one given in Aitchison and Dunsmore (1975), except that the levels of concentration of antibiotic are slightly different. The doctors presenting the data were interested in (i) investigating the relationship between log concentration of Tobramycin and clearance diameter and (ii) examining the feasibility of using this biological assay data to estimate the concentration of Tobramycin in patient's blood. The calibration data set used in Chapters 3, 4 and 5 consisted of 67 observations with the level of concentration of Tobramycin ranging from 0.6 μg/ml to 10 μg/ml.

A1.3 THE TEETH DATA SETS

Two different groups of teeth data are considered.

Group A. The teeth data sets used in Chapters 3, 4 and 5 relate to the length of transparent root dentine (ITTM) measured on 153 non-carious intact teeth extracted from patients at the Glasgow Dental Hospital. A total of 43 patients contributed the teeth, their ages ranging from 16.1 years to 65.6 years. The 153 teeth were classified into six types: upper central, upper lateral, upper canine, lower central, lower lateral and lower canine (the upper teeth being on the upper jaw and the lower teeth being on the lower jaw). The number of teeth obtained from each patient ranged from 1 to 12. The position in the mouth, right or left, was also recorded. There is considerable similarity between right and left teeth of the same type, so on dental grounds alone, right and left differences were ignored. The roots of all teeth contain root dentine. Part of the root dentine is transparent and the remainder is opaque. Fig. A1.1 shows a photograph of such a root and the regions of transparent and opaque root dentine are marked. The amount of transparent root dentine is known to be related to age. Although the photograph is of a ground section of a root, the measurements of transparent root dentine which make up the six data sets considered, were made on the intact root *before* it was sectioned. The ITTM reading was obtained by measuring the *length* of transparent root dentine, to the nearest 0.1mm, on opposite sides of the intact root using a vernier caliper and then averaging these two measurements.

Fig. A1.1 Ground section of a root (Tooth 1)



Group B. This data set was used in Chapter 6 only. It consisted of measurements of the area of transparent root dentine (TRA) as a proportion of the total root dentine (TRT) for upper canine teeth only. As tooth size and hence root size varies considerably, it was thought better to measure the *proportion* of transparent dentine rather than the *absolute length* of transparent dentine as used in Group A data sets. The procedure of obtaining the measurements TRA/TRT consisted of obtaining a ground section of the root, photographing at about 5 x magnification, then preparing contact prints of each root section and placing them on the bit-pad connected to a computer. The perimeter of each required area, TRA or TRT, was traced with the light spot of a probe. For each root, three independent readings of TRA and TRT were made. Fig. A1.2 shows a picture of a different root from that photographed for Fig. A1.1. The region of the light shading shows the area of transparent dentine (TRA) whilst the area with heavy mottling shows the area of opaque dentine. The two areas combined together give the total area of root dentine (TRT). If one compares Fig. A1.1 and Fig. A1.2, one will see that the pattern of transparent root dentine is very different in the two roots. One would imagine that trying to obtain the measurement ITTM for a root like that shown in Fig. A1.2 would be fraught with difficulties.

A total of 14 patients contributed 17 upper canine teeth so the calibration data set used in Chapter 6 consisted of 51 observations.

REFERENCES

Aitchison, J. & Dunsmore, I.R. (1975). *Statistical Prediction Analysis.* Cambridge University Press.

Preece, M.A. & Baines, M.J. (1978). A new family of mathematical models describing the human growth curve. *Anns. Hum. Biol.*, **5**, 1-24.

Ratkowsky, D.R. (1983). *Nonlinear regression models.* Marcel Dekker.

Fig. A1.2 Ground section of a root (Tooth 2)